WordsEye: An automatic text-to-scene conversion system

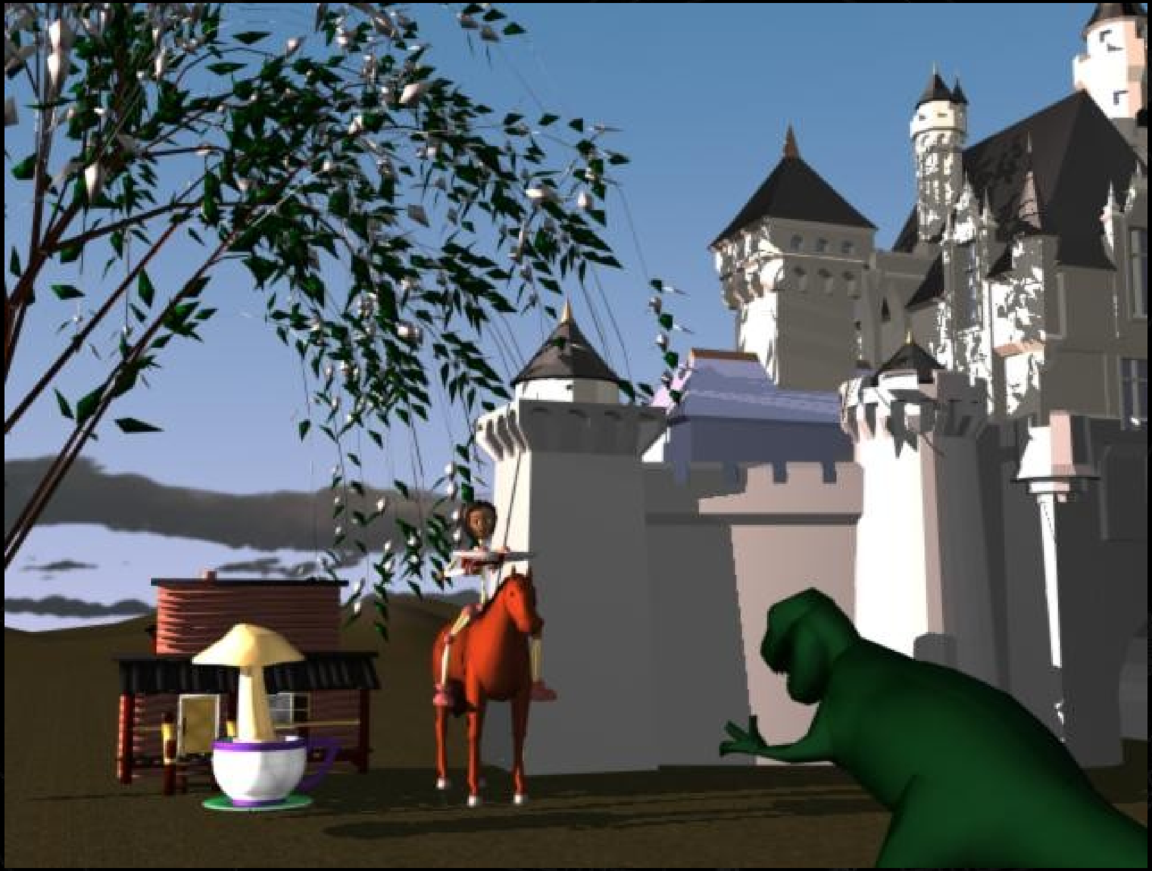
Bob Coyne, Richard Sproat
AT&T Labs - Research

Web site: www.research.att.com/projects/wordseye

WordsEye Introduction

- **Visualize the meaning of language**
  - Flexible syntax, semantics, reference
- **Effortless, immediate…just describe it**
  - No skill or training required
  - No interface to get in the way
  - Paint a picture with words.
- **Enable novel applications**

Mary uses the crossbow. She rides the horse by the store. The store is under the large willow. The small allosaurus is in front of the horse. The dinosaur faces Mary. The gigantic teacup is in front of the store. The gigantic mushroom is in the teacup. The castle is to the right of the store.



AT&T Labs - Research

Related Work

- Adorni, Di Manzo, Giunchiglia, 1984
- Put: Clay and Wilhelms, 1996
- PAR: Badler et al., 2000
- CarSim: Dupuy et al., 2000
- SHRDLU: Winograd, 1972

Implementation

- 1 1/2 years development
  - Completing initial version
- Written in Common Lisp on Windows NT
  - Uses Mirai animation system
- Parser/Tagger in C on Linux.
- Viewpoint 3D model library

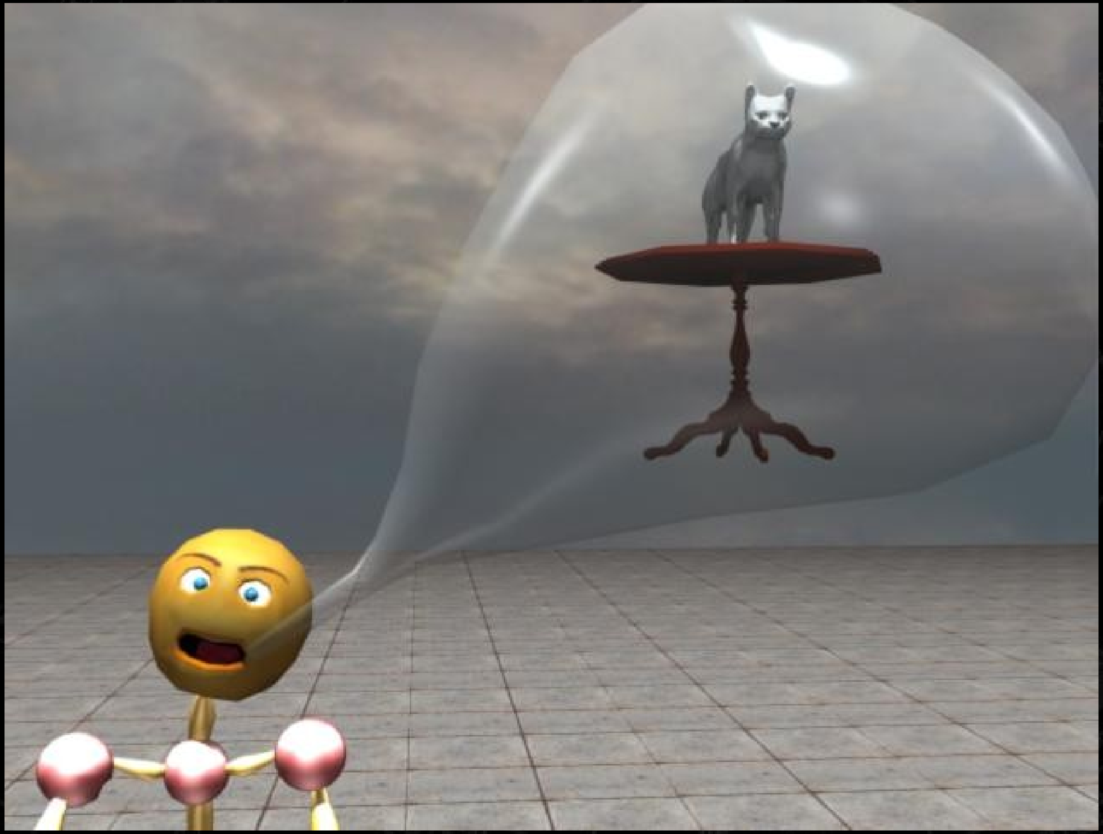
WordsEye Overview

- **Linguistic Analysis**
  - Parsing, semantic representation
- **Interpretation**
  - Add implicit objects, relations
  - Resolve references
- **Depiction**
  - Database of 3D objects, poses
  - Depiction rules generate graphical *depictors*
  - Apply depictors to create scene

Linguistic Analysis

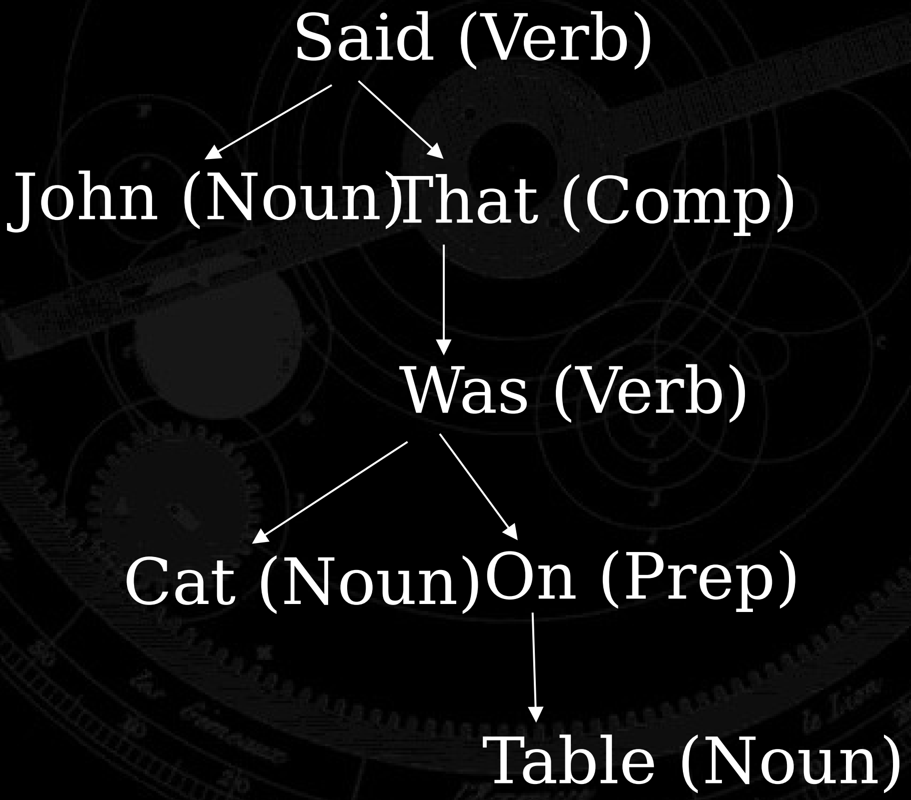
- Tag part-of-speech (Church, 1988)
- Parse (Collins, 1999)
- Generate semantic representation
  - Semantic functions for verbs and prepositions
  - WordNet (Fellbaum, 1998) for nouns
- Anaphora resolution

Depiction: *John said that the cat is on the table.*



AT&T Labs - Research

Parse tree for: *John said that the cat was on the table.*



Said (Verb)

John (Noun)

That (Comp)

Was (Verb)

Cat (Noun)

On (Prep)

Table (Noun)

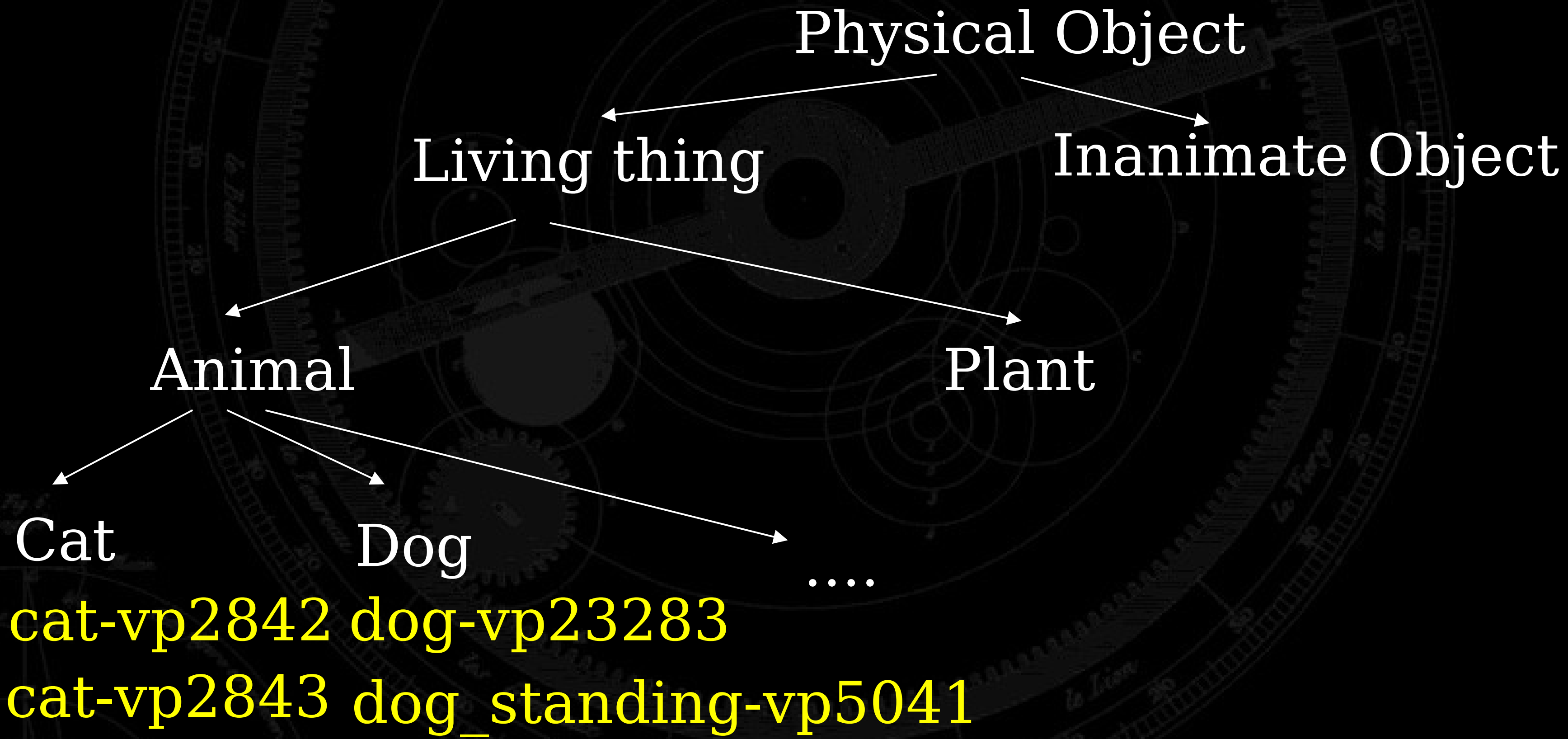
Semantic Function for *Say*

```
(semantics :genus “say”
   (verb-frame
      :name say-believe-that-s-frame
      :required (subject that-s-object)
      :optional (actionlocation actiontime))
   (verb-frame
      :name say-believe-s-frame
       :required (subject s-object)
      :optional (actionlocation actiontime)))
```

Semantic Function for *On*

```
(semantics :genus “on”
  (pframe “Time” (:any :temporal))
  (pframe “Attachment” (:any :attachment))
  (pframe “Riding-large-vehicle” (:any :large-vehicle))
  (pframe “Contiguous-with-wall” (:houseware :wall))
  (pframe “On-geographical-area” (:any :geo-area))
  (pframe “Contiguous-with” (:building :geo-entity))
  (pframe “Default:location-support-entity” (:any
    :any)))
```

Nouns: Extended WordNet Hierarchy



Physical Object

Living thing

Inanimate Object

Animal

Plant

Cat

Dog

....

cat-vp2842 dog-vp23283

cat-vp2843 dog_standing-vp5041

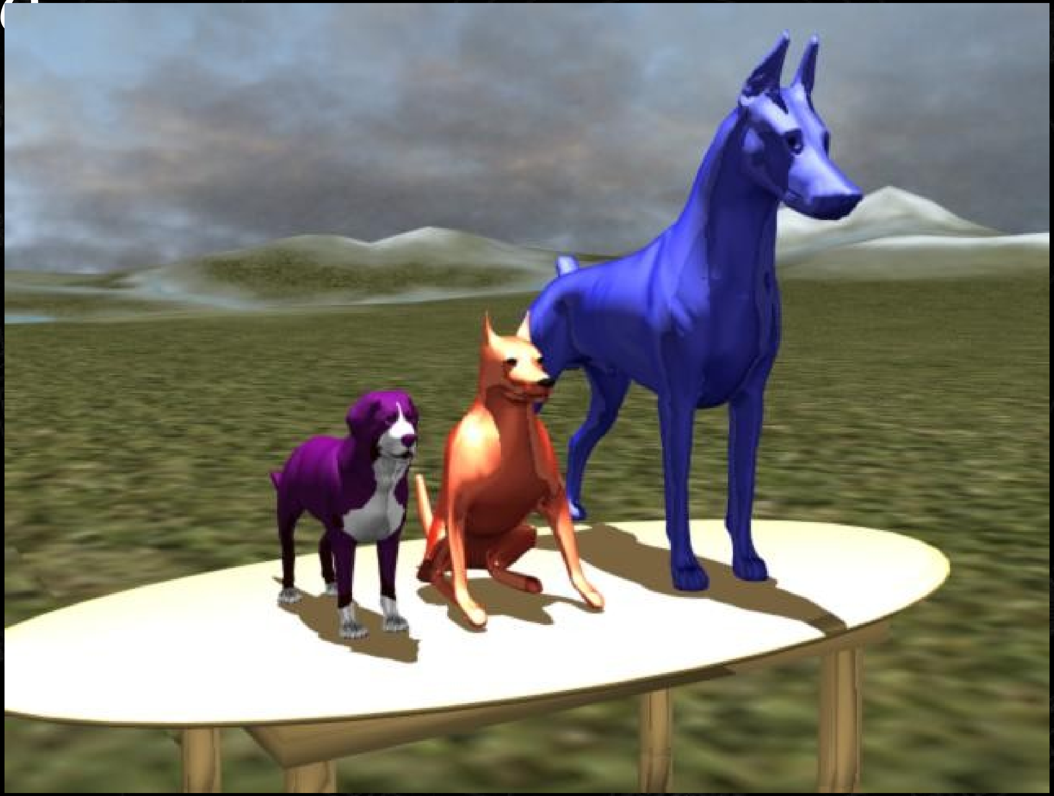
Semantic Representation for: *John said that the blue cat was on the table.*

1. Object: “mr-happy” (John)
2. Object: “cat-vp39798” (cat)
3. Object: “table-coffee-vp6204” (table)
4. Action: “say” :subject <element 1>
   :direct-object <elements 2,3,5,6> :tense “PAST”
5. Attribute: “blue” :object <element 2>
6. Spatial-Relation “on” :figure <element 2>
   :ground <element 3>

Interpretation

- **Interpret semantic representation**
  - Answer *Who? What? When? Where? How?*
  - Disambiguate/canonicalize relations and actions
  - Identify implicit objects
  - Reference resolution

Indexical Reference: *Three dogs are on the table. The first dog is blue. The first dog is 5 feet tall. The second dog is red. The third dog is purple.*

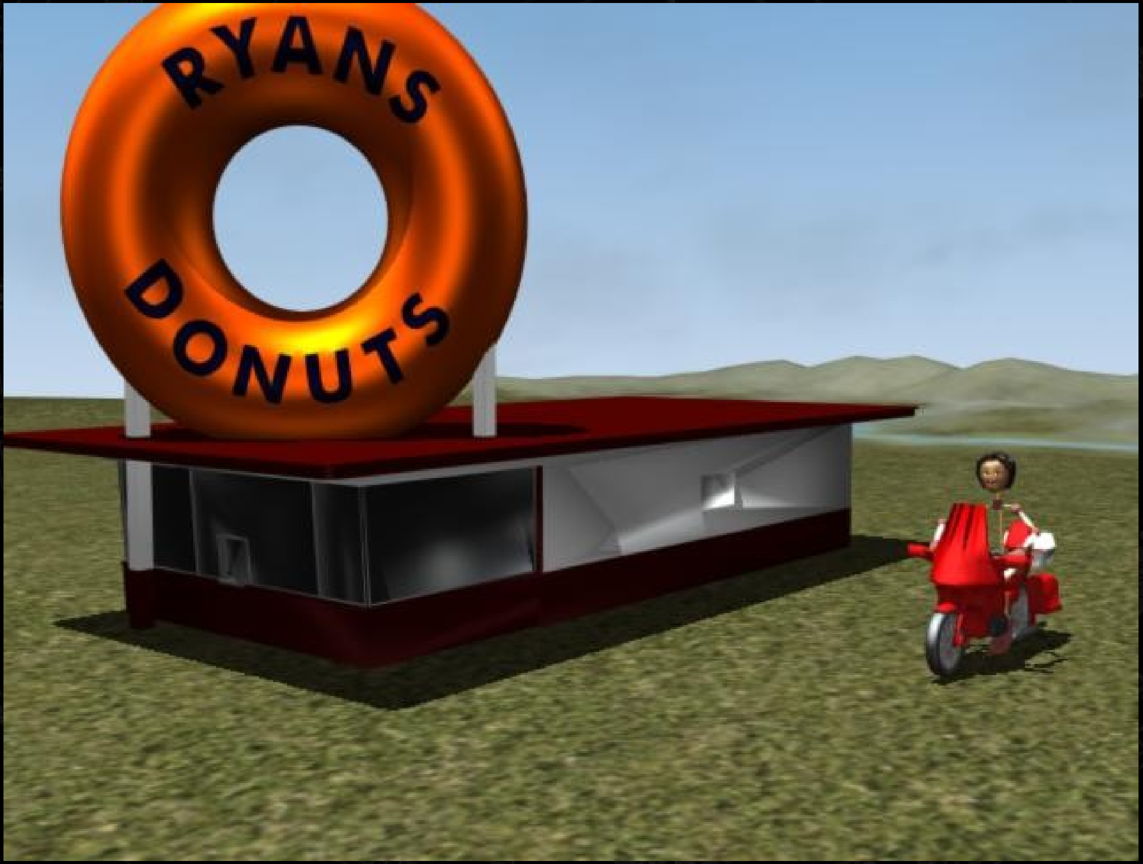


AT&T Labs - Research

Implicit objects & references

- ***Mary rode* by the store. *Her motorcycle* was red.**
  - Verb resolution: Identify implicit *vehicle*
    - Functional properties of objects
  - Reference
    - *Motorcycle* matches the *vehicle*
    - *Her* matches with *Mary*

Implicit Reference: *Mary rode by the store. Her motorcycle was red.*



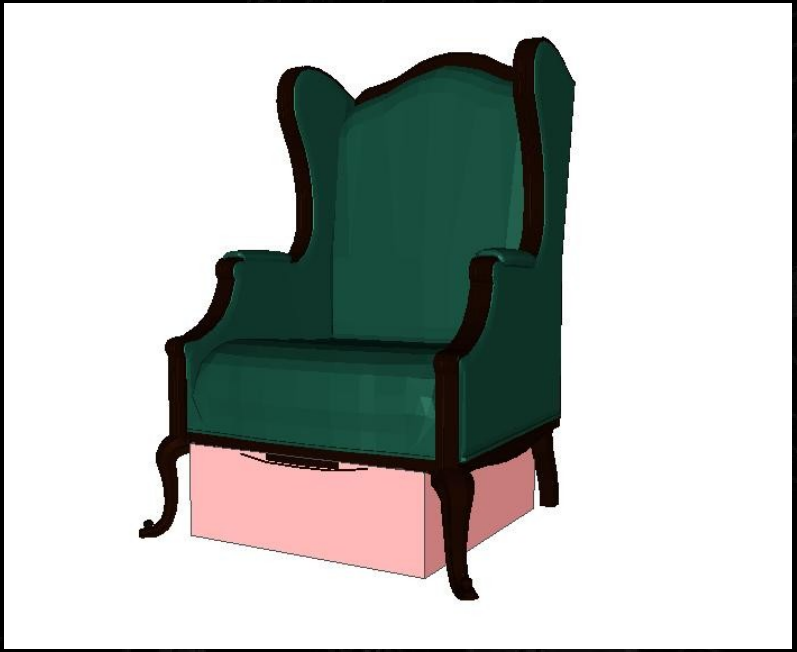
Depiction

- **3D object database**
- **Graphical operations (depictors)**
  - Spatial relations
  - Attributes
  - Posing
  - Shape/Topology changes
- **Depiction process**

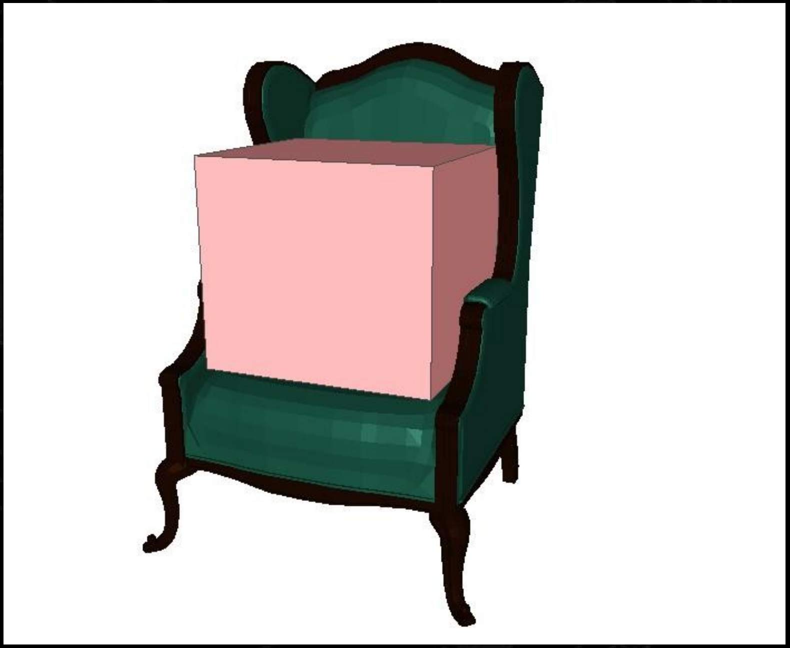
3D Object Database

- **2,000+ 3D polygonal objects (Viewpoint+)**
- **Augmented with:**
  - Skeletons
  - Default size, orientation
  - Functional properties (*vehicle, weapon*,..)
  - Placement/attribute conventions
  - Spatial tags (top surface, base, cup, push handle, wall, stem, enclosure)

Spatial Tags



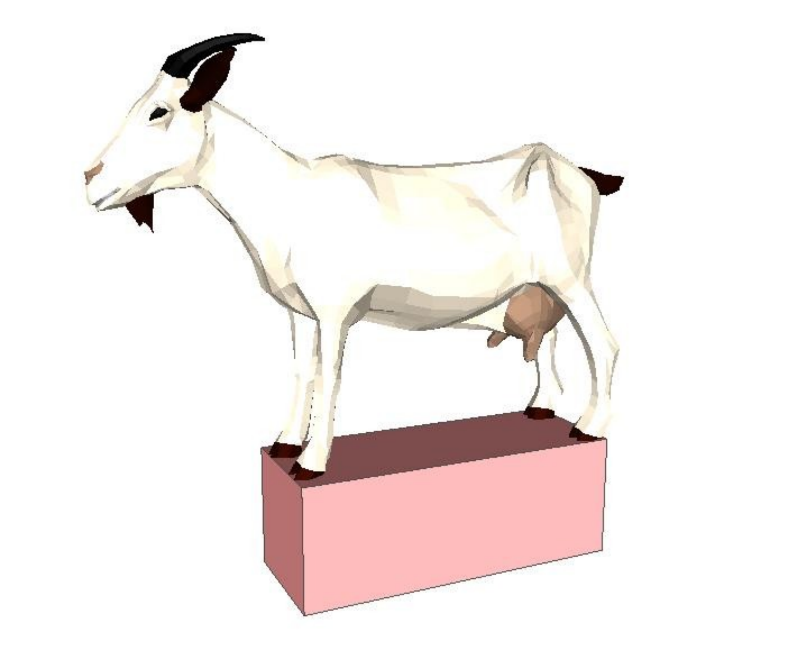
Canopy (under, beneath)



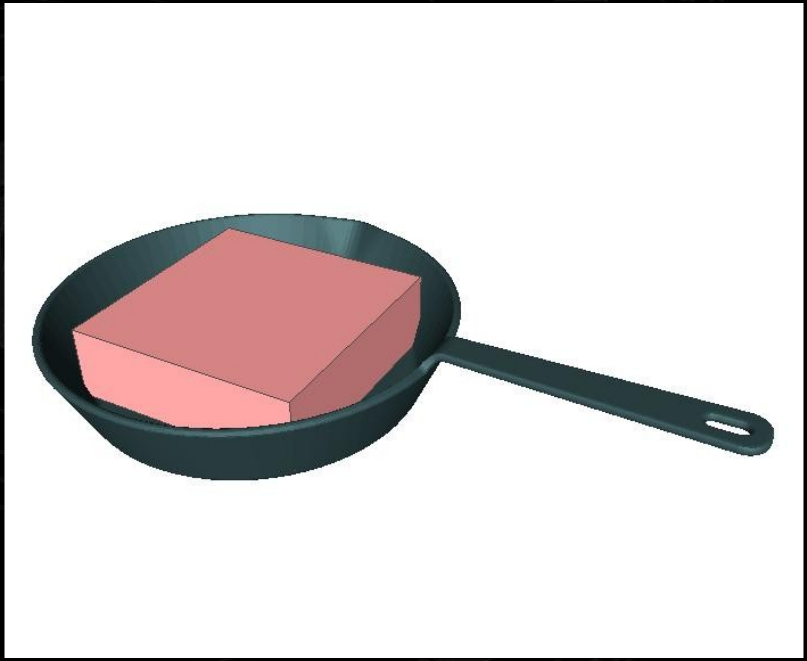
Top Surface (on, in)

AT&T Labs - Research

Spatial Tags



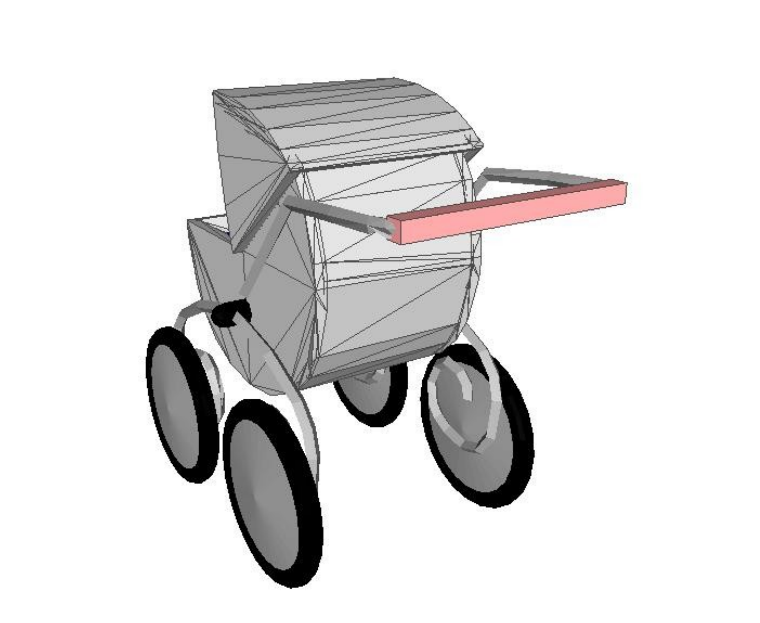
Base (under, below, on)



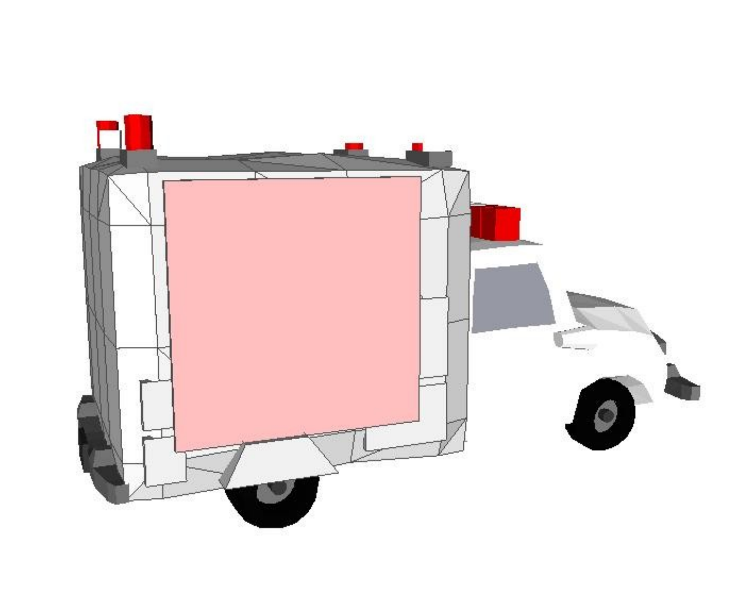
Cup (in, on)

AT&T Labs - Research

Spatial Tags

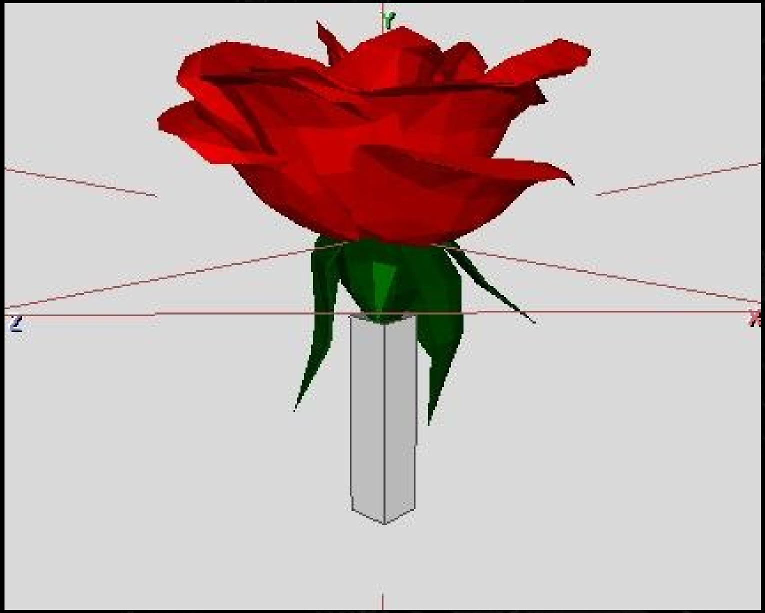


Push Handle (actions)

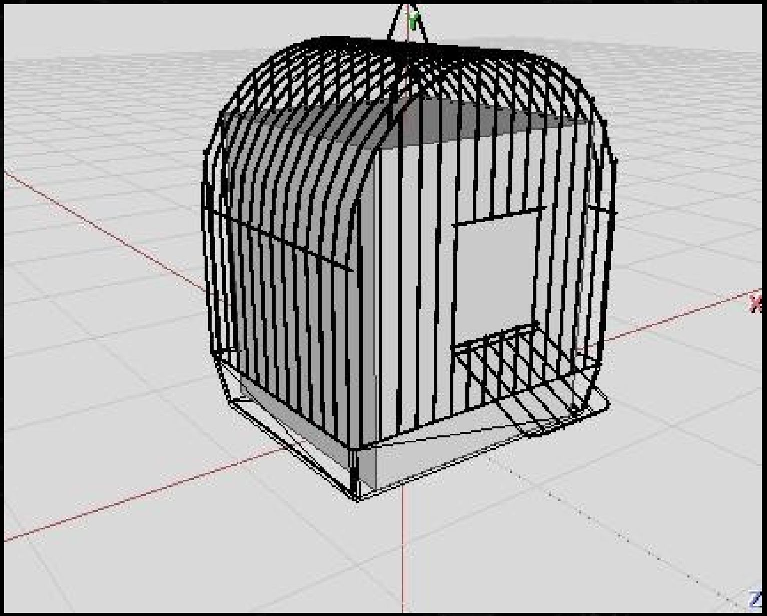


Wall (on, against)

Spatial Tags



Stem (in)

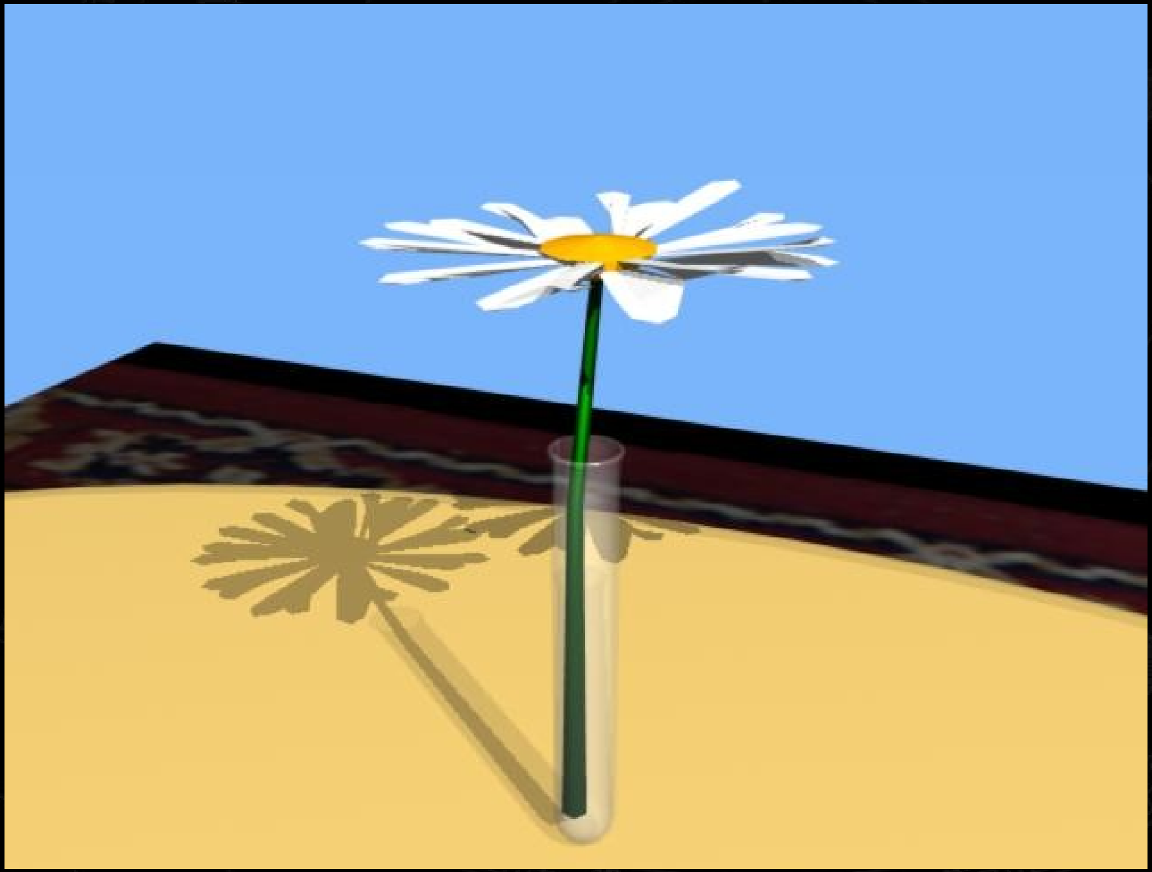


Enclosure (in)

Spatial Relations

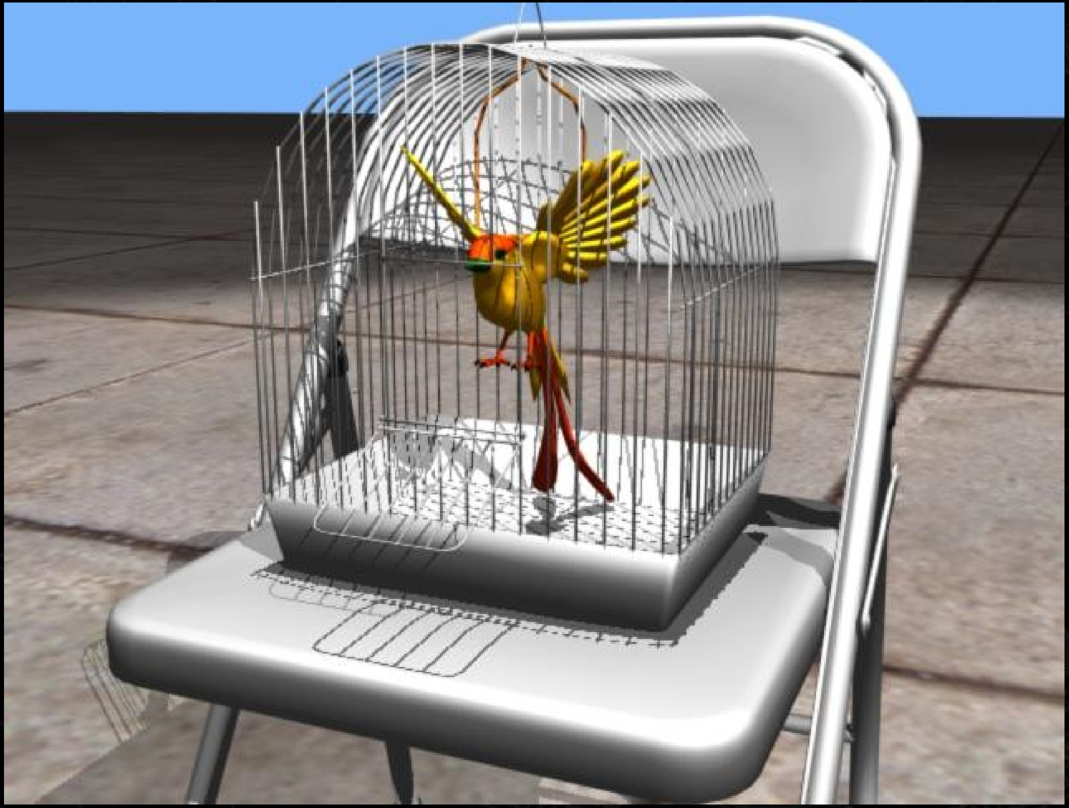
- **Relative positions**
  - On, under, in, below, off, onto, over, above, ….
  - Distance
- **Subregion positioning**
  - Left, middle, corner,right, center, top, front, back
- **Orientation**
  - facing (***object***, left, right, front, back, east, west,…)
- **Time-of-day relations**

Stem in Cup: *The daisy is in the test tube.*



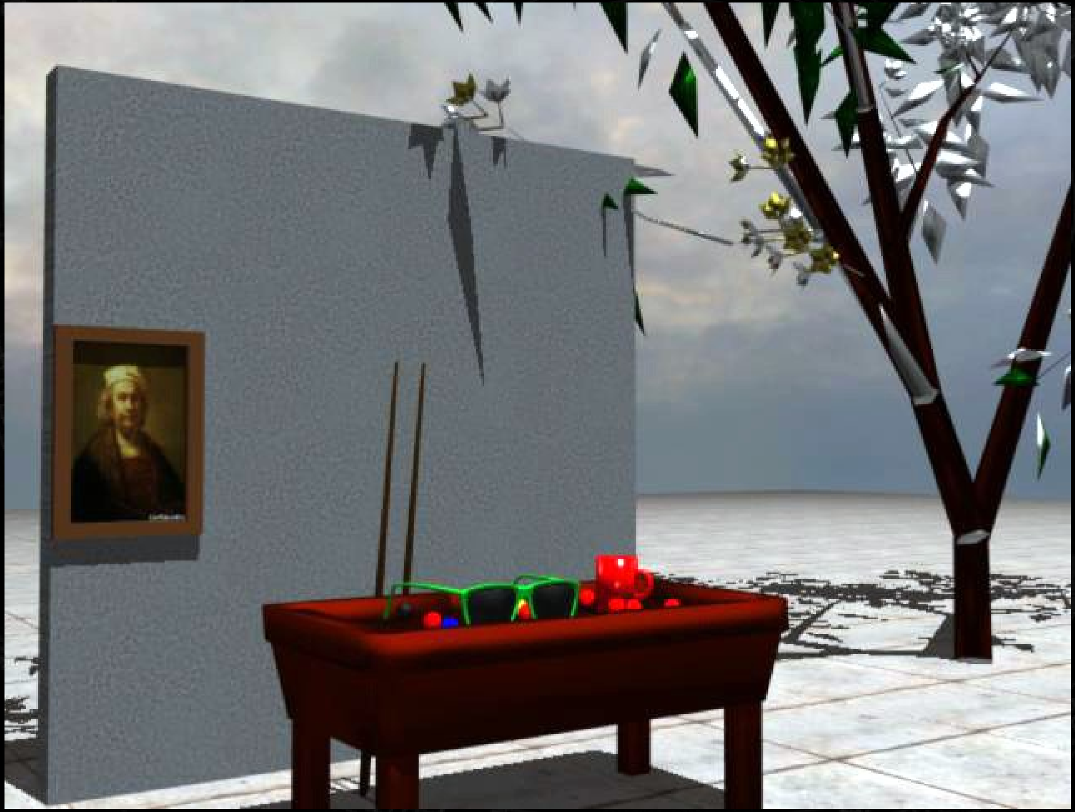
AT&T Labs - Research

Enclosure and top surface: *The bird is in the bird cage. The bird cage is on the chair.*

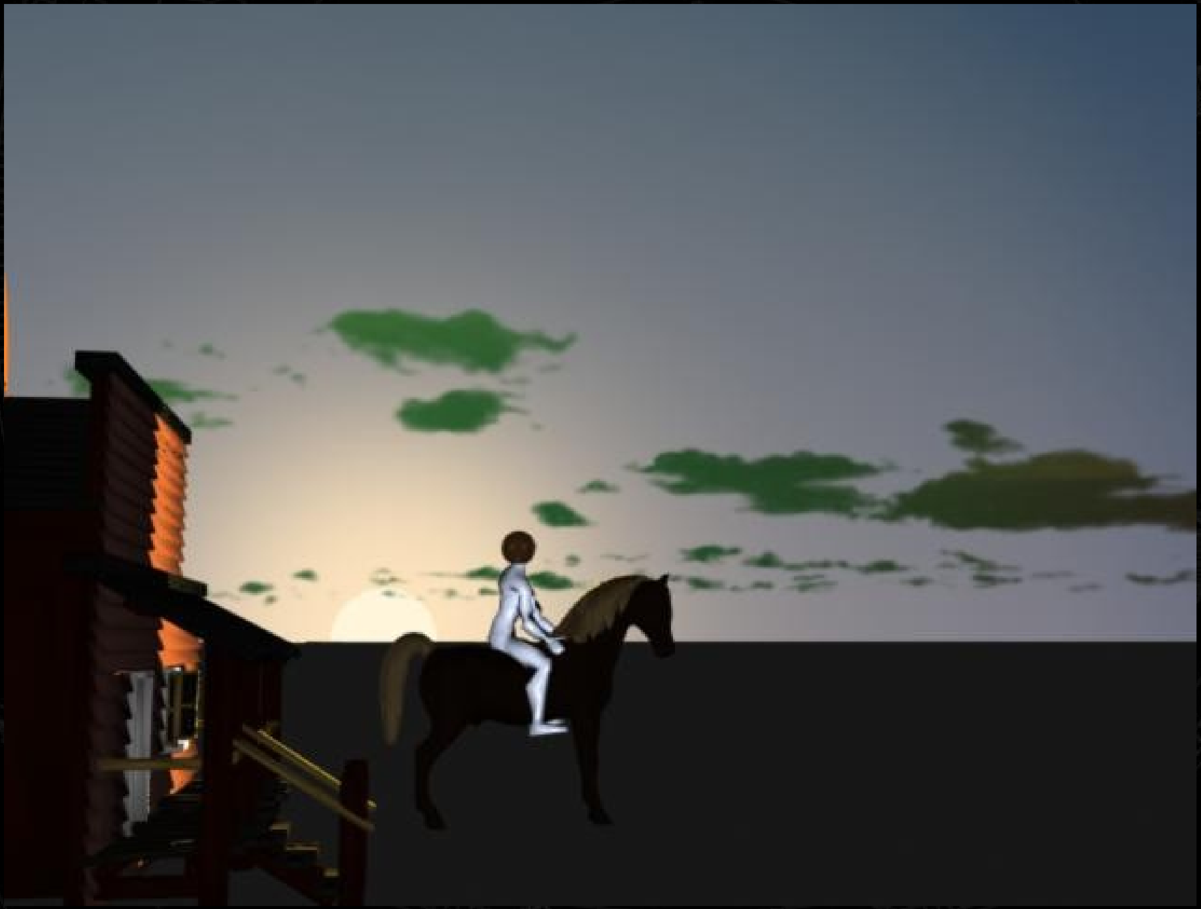


AT&T Labs - Research

Image DB, placement: *The large red coffee mug and huge sunglasses are on the table. The table is in front of the wall. The picture of Rembrandt is on the left of the wall. The wall is under the cherry tree.*



Time relation: *At 7 a.m., John rides the horse…*

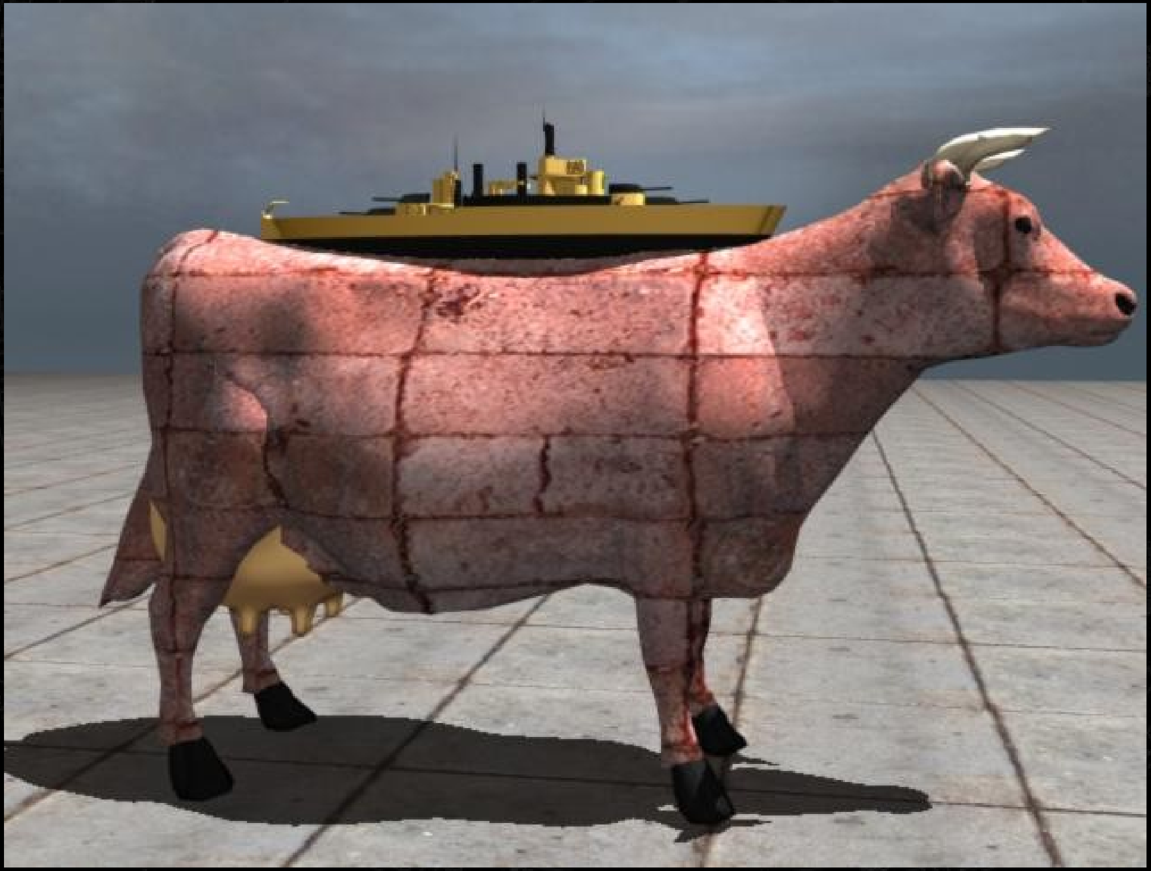


AT&T Labs - Research

Attributes

- **Size**
  - height, width, depth
  - Aspect ratio (flat, wide, thin,...)
- **Surface attributes**
  - Texture database
  - Color, Texture, Opacity
  - Applied to objects or textures themselves

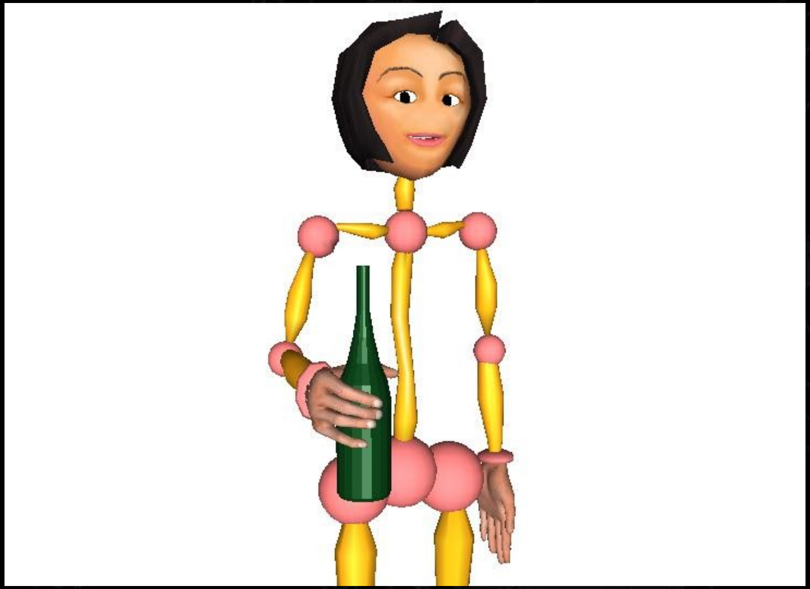
Attributes: *The orange battleship is on the brick cow. The battleship is 3 feet long.*



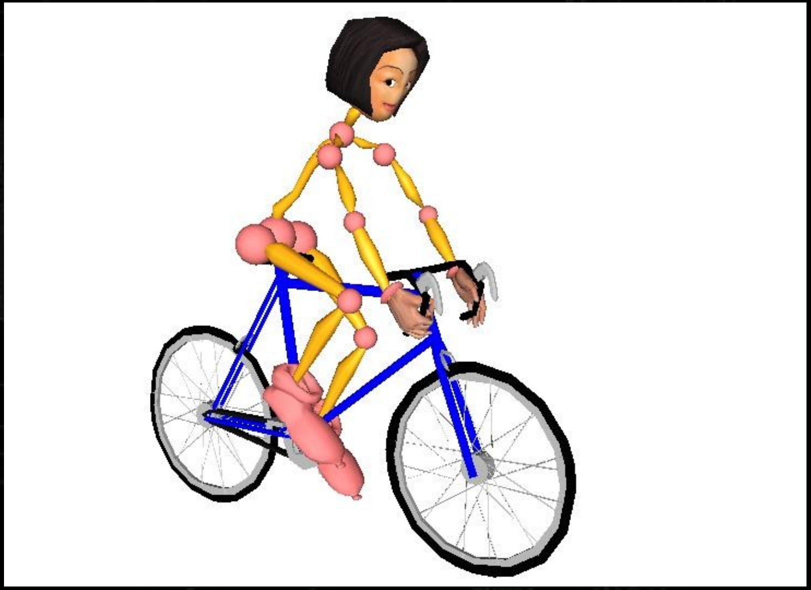
Poses

- **Represent actions**
- **Database of 500+ human poses**
  - Grips
  - Usage (specialized/generic)
  - Standalone
- **Merge poses (upper/lower body, hands)**
- **Dynamic posing/IK**

Poses



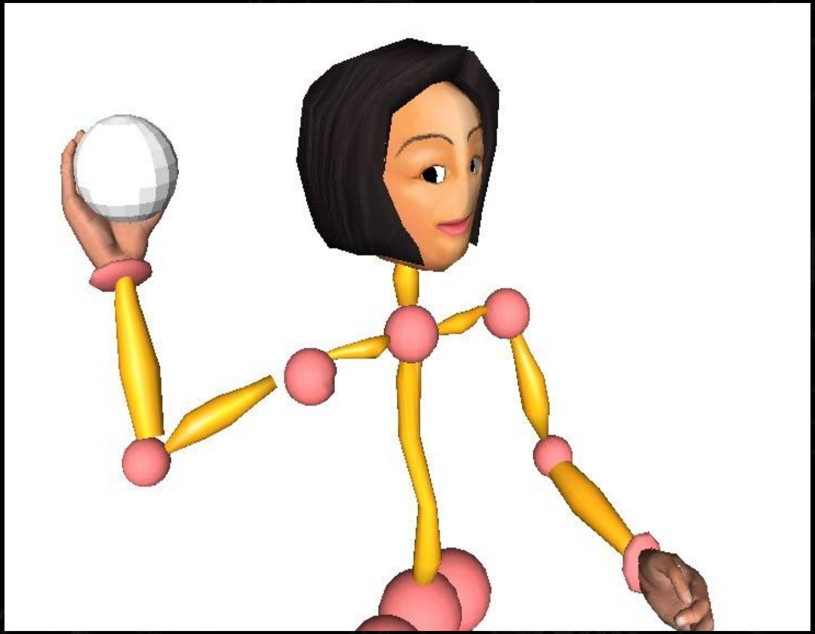
Grip wine_bottle-bc0014



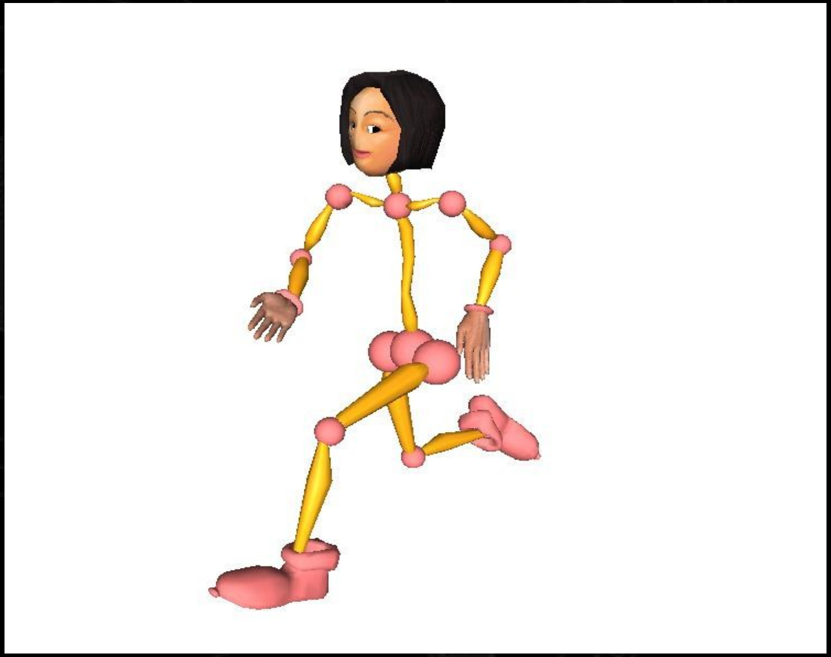
Use bicycle_10-speed-vp8300

AT&T Labs - Research

Poses



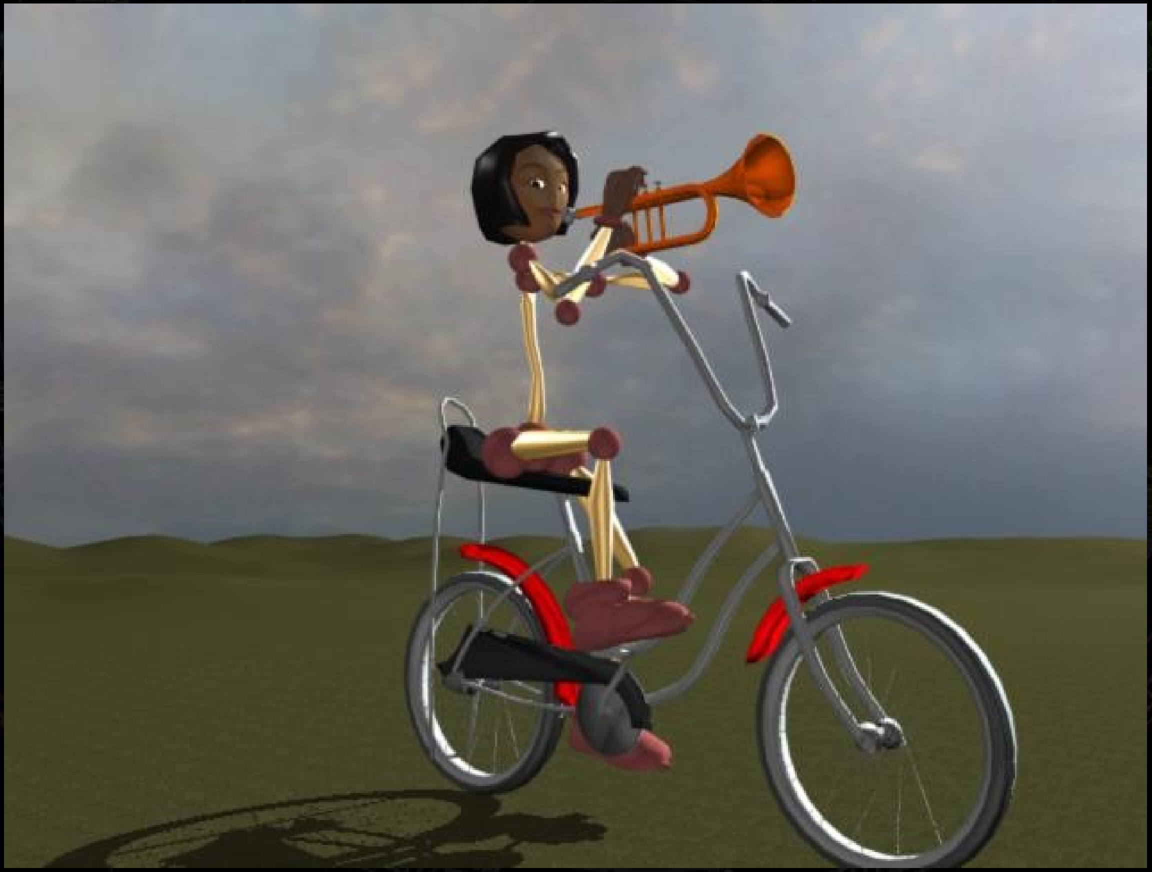
Throw "round object"



Run

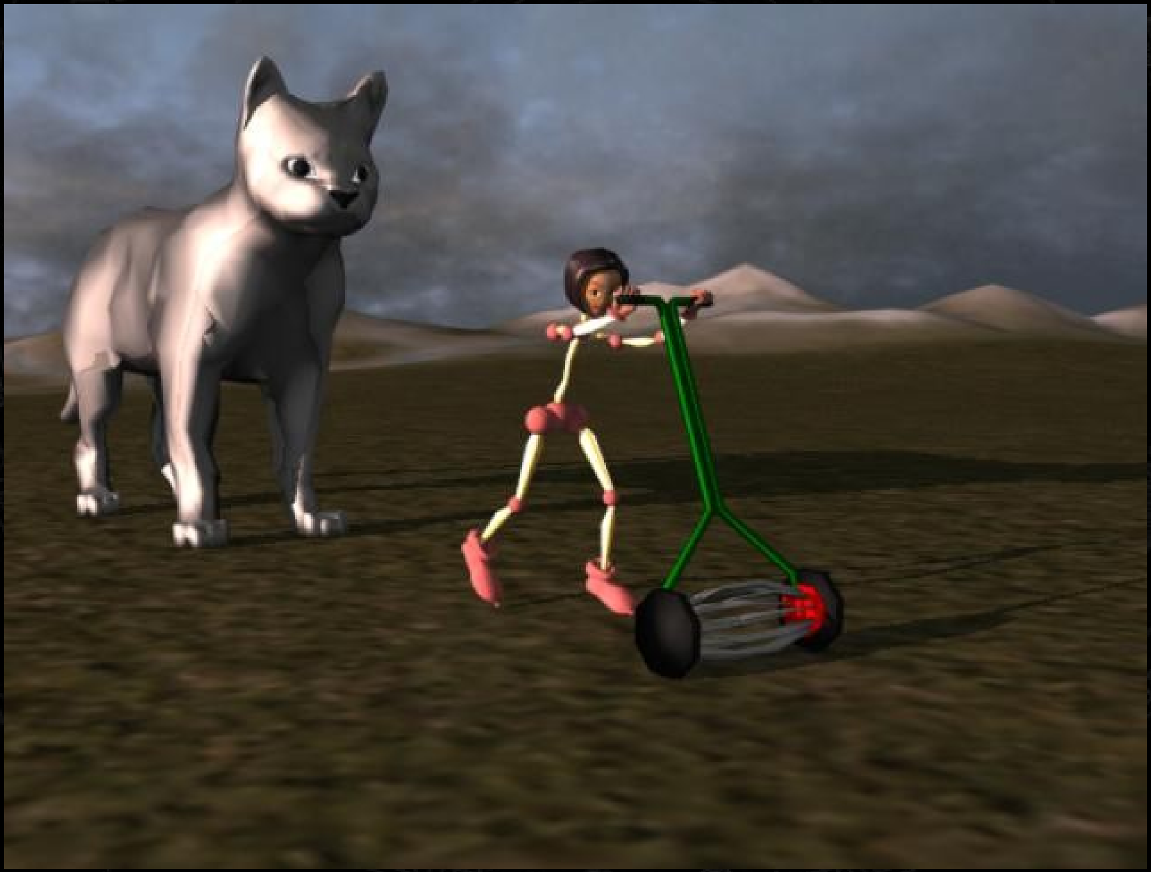
AT&T Labs - Research

Combined poses: *Mary rides the bicycle. She plays the trumpet.*



AT&T Labs - Research

Inverse Kinematics (IK): *Mary pushes the lawn mower. The lawnmower is 5 feet tall. The cat is 5 feet behind Mary. The cat is 10 feet tall.*

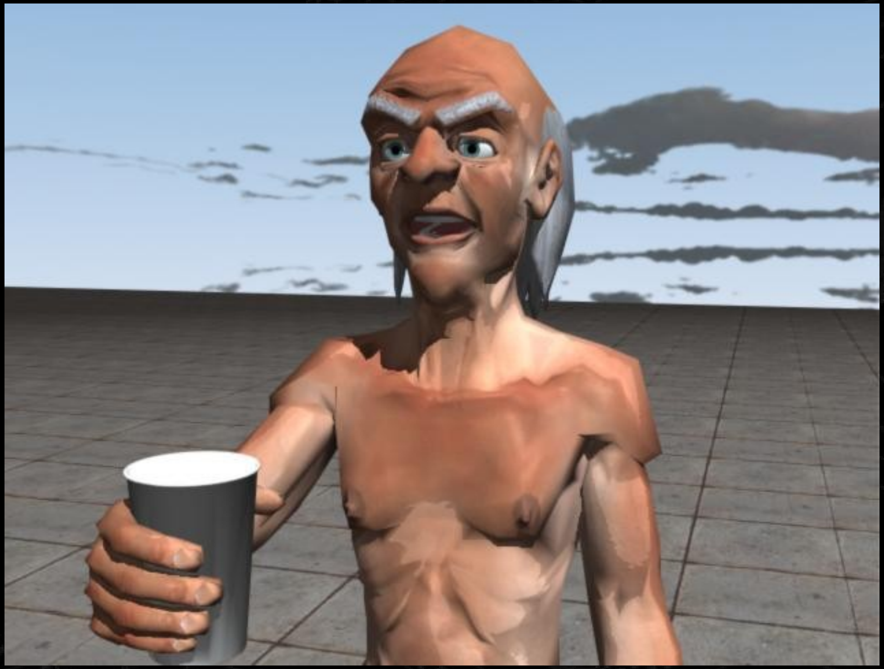


AT&T Labs - Research

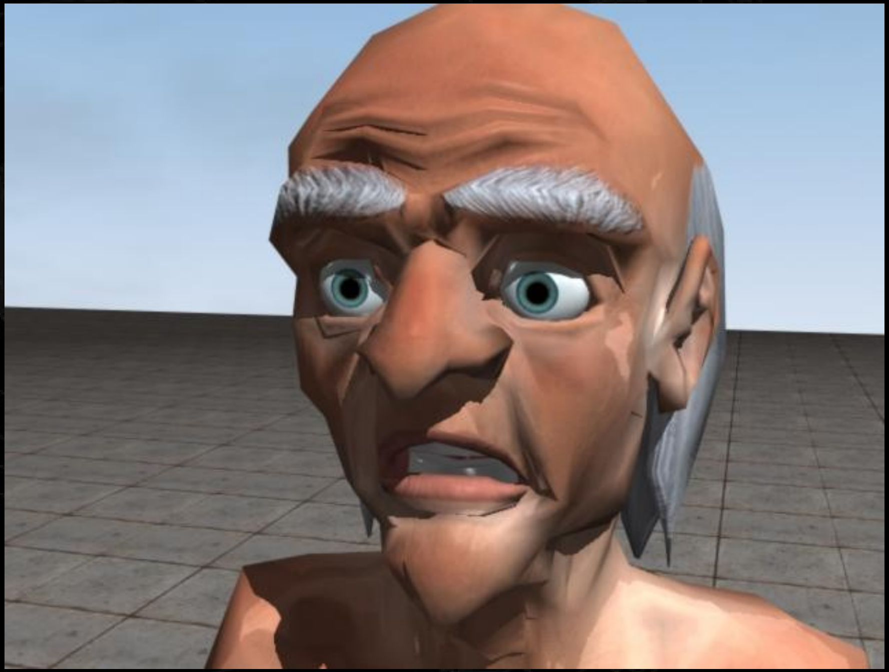
Shape Changes

- **Deformations**
  - Facial expressions
    - Happy, angry, sad, confused,…mixtures
    - Combined with poses
- **Topological changes**
    - Slicing

Facial Expressions



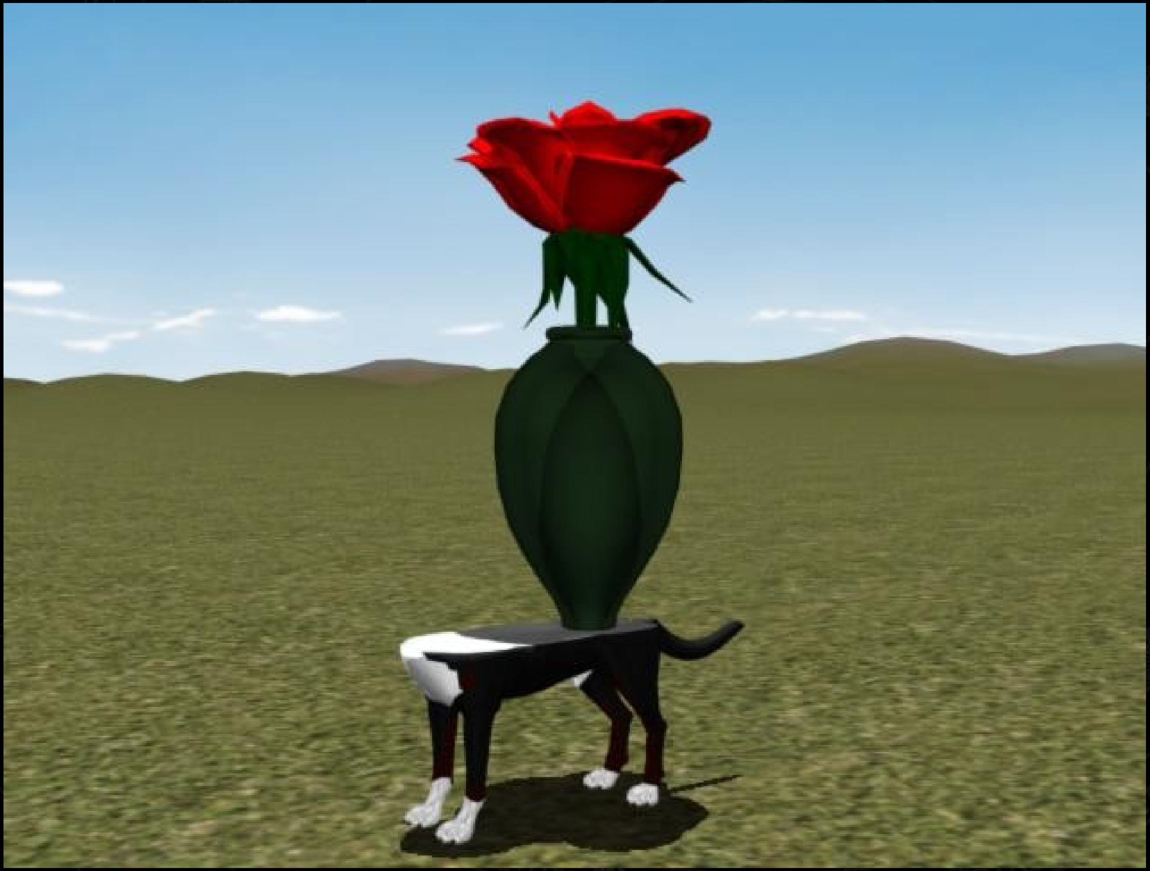
Edward holds the cup.
He is happy.



Edward is shocked.

AT&T Labs - Research

The rose is in the vase. The vase is on the half dog.



AT&T Labs - Research

Depiction Process

- **Given a semantic representation**
  - Generate *depictors* (specs for graphical operations)
  - Modify depictors to handle implicit and conflicting constraints.
  - Generate 3d scene from depictors
  - Add environment, lights, camera
  - Render scene

Example: Generate depictors for *kick*

Case1: *No path or recipient; Direct object is large*

- Pose: Actor in *kick pose*
- Position: Actor directly behind direct object
- Orientation: Actor facing direct object

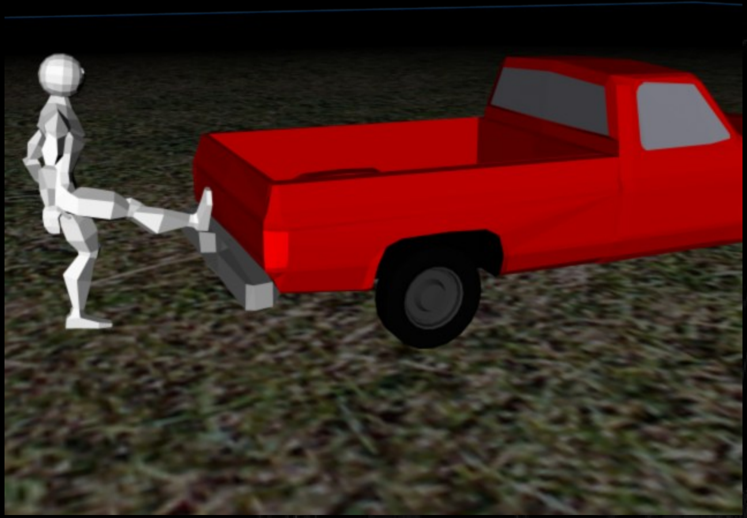
Case2: *No path or recipient; Direct object is small*

- Pose: Actor in *kick* pose
- Position: Direct object above *foot*

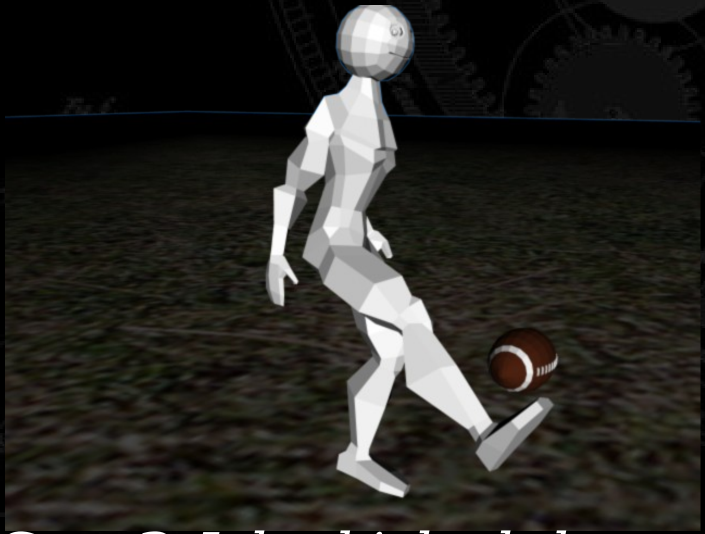
Case3: *Path and Recipient*

- ***Pose+relations…(some tentative)***

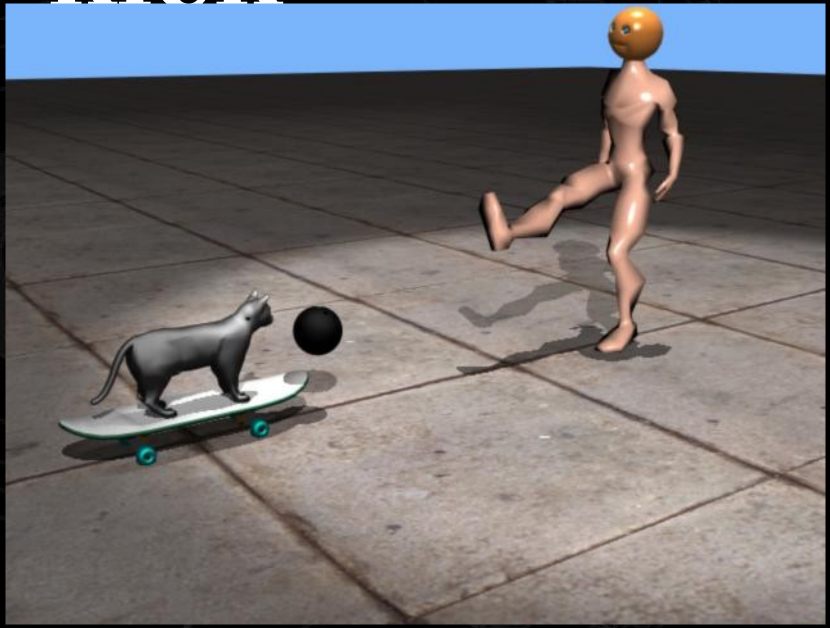
Varieties of kick



Case1: *John kicked the pickup truck*



Case2:*John kicked the football*



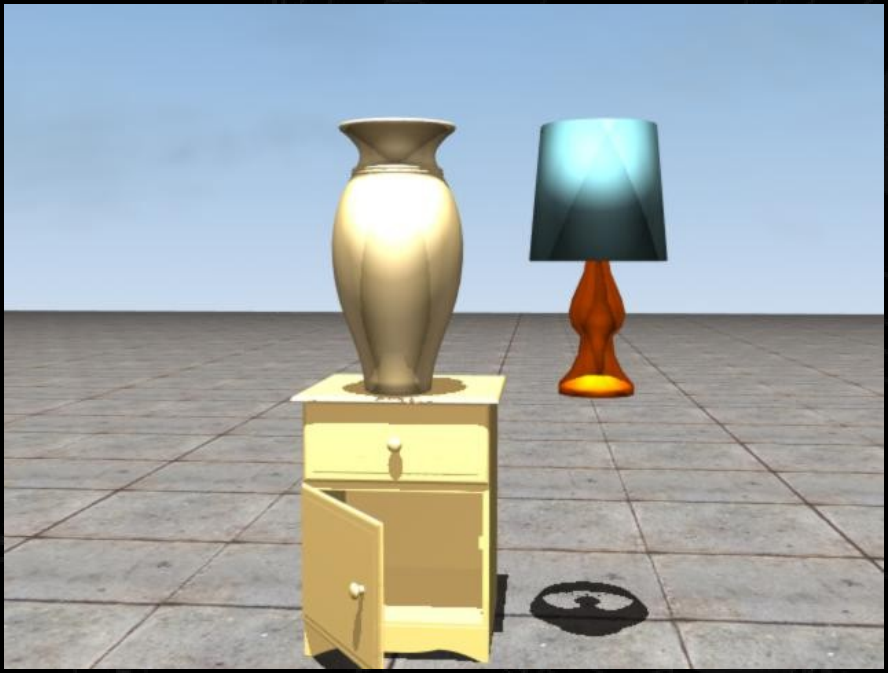
Case3: *John kicked the ball to the cat on the skateboard*

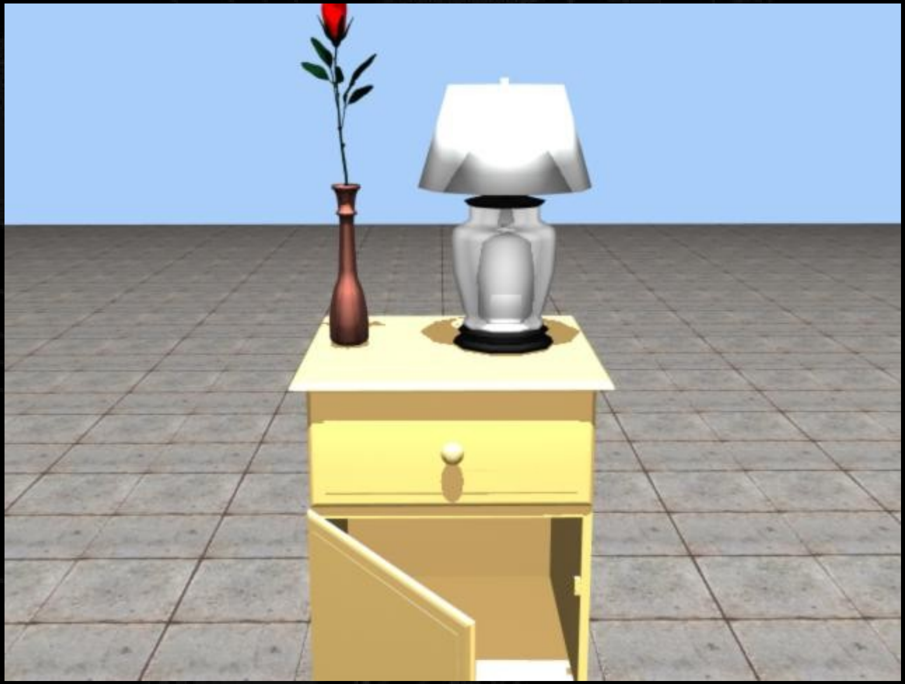
AT&T Labs - Research

Example: Modify Depictors by adding Implicit Constraints

- If A is next to B and not on a surface, put A on same surface as B. (unless A is airborne)
  - e.g. *The vase is on the table. The lamp is next to the vase.*
- If A and B are on same surface and not laterally constrained, put A next-to B.
  - e.g. *The dog and cat are on the table.*

Implicit Constraint. *The vase is on the nightstand. The lamp is next to the vase.*





AT&T Labs - Research

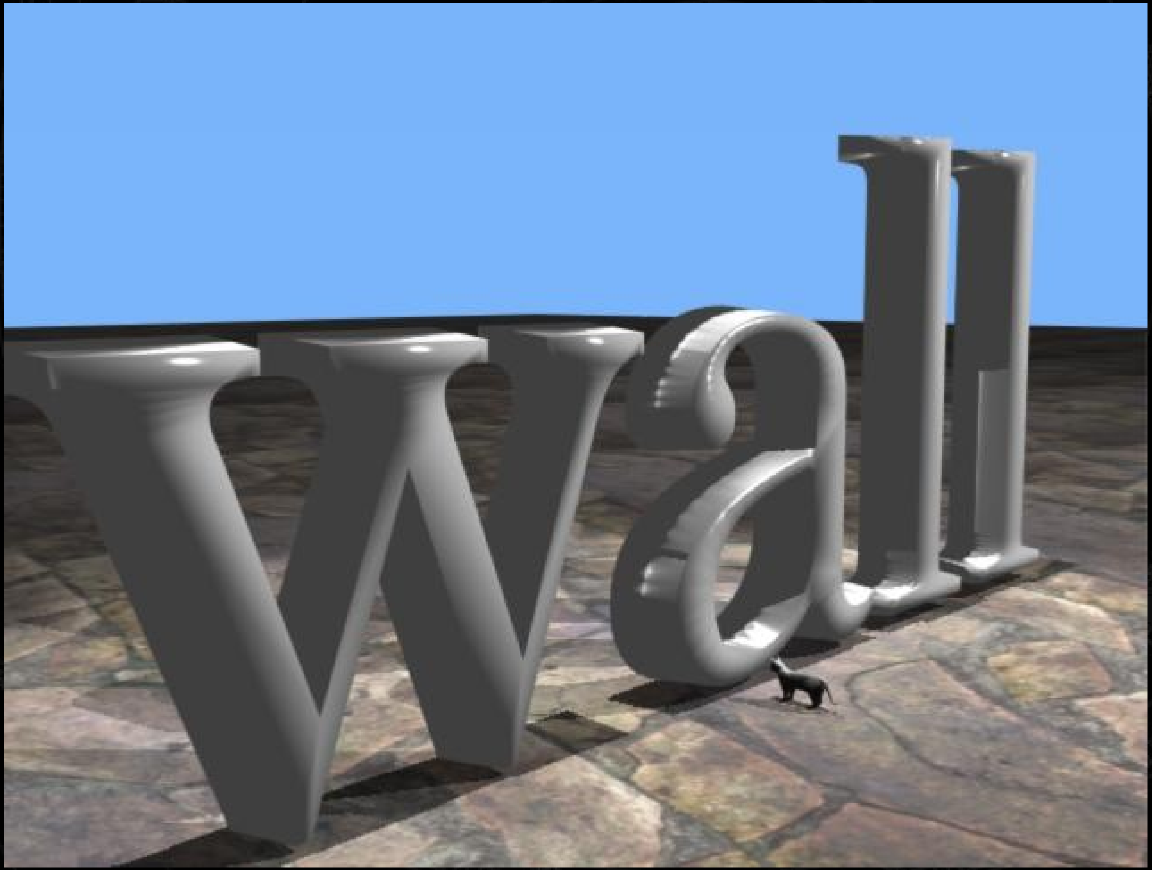
Generate Scene from Depictors

- **For objects on each surface:**
  - Set initial size, orientation, shape, color
  - Apply pose/shape changes. Attach held objects
  - Move objects, maintaining constraints
  - Apply relative orientations
  - Apply dynamic operations (IK, objects on paths)

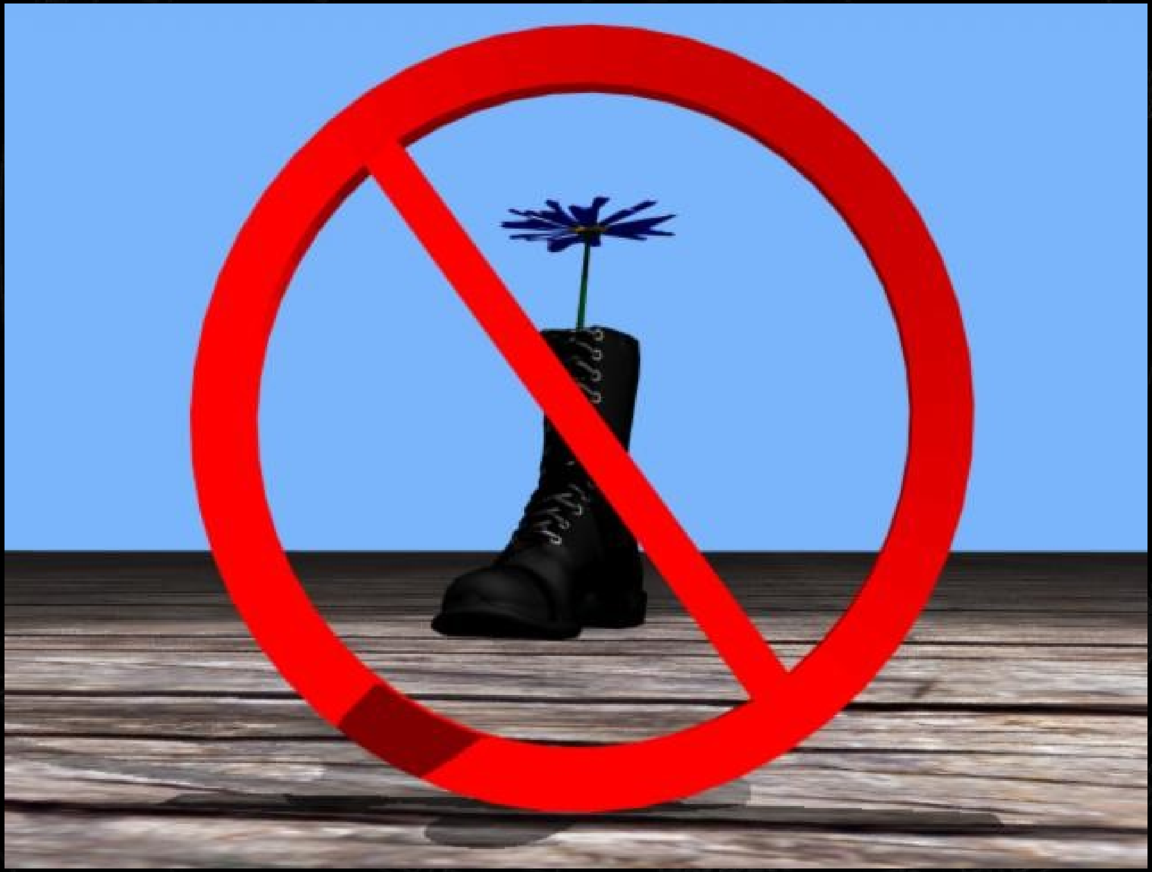
Figurative & Metaphorical Depiction

- Textualization
- Conventional Icons and emblems
- Literalization
- Characterization
- Personification
- Functionalization

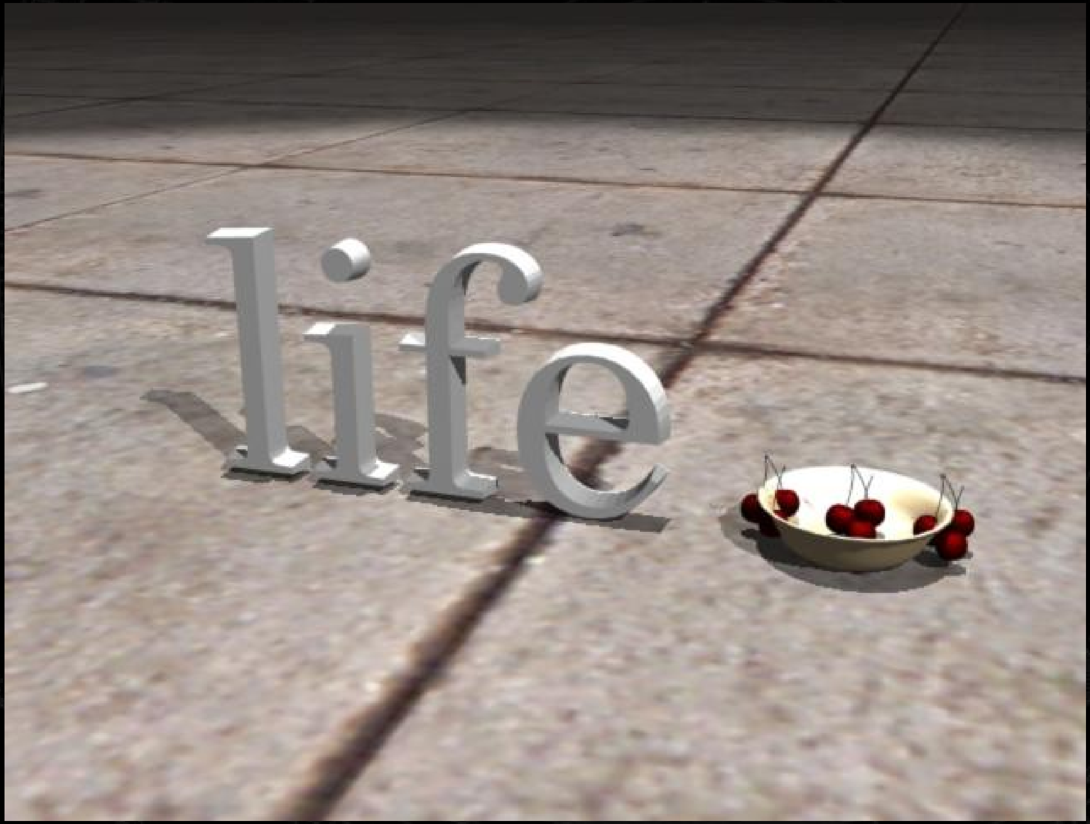
Textualization: *The cat is facing the wall.*



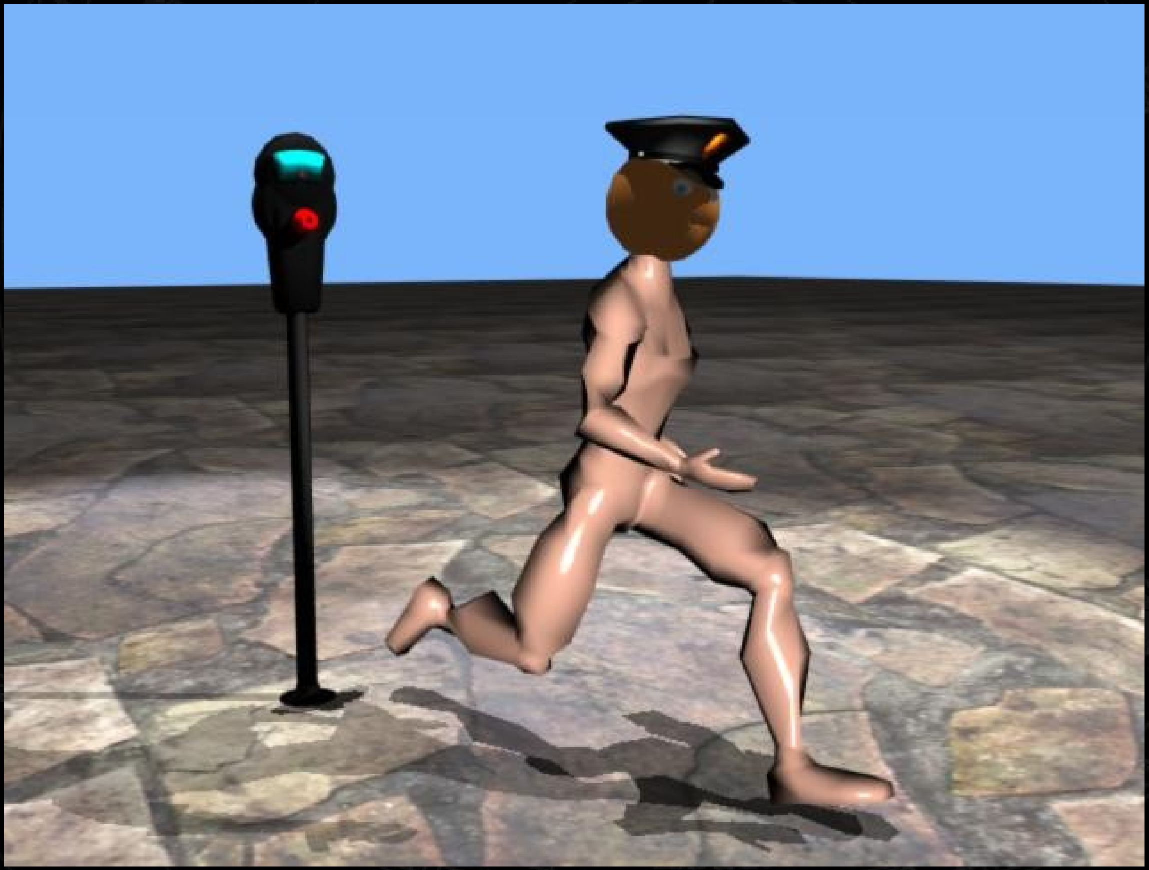
Conventional Icons: *The blue daisy is not in the army boot.*



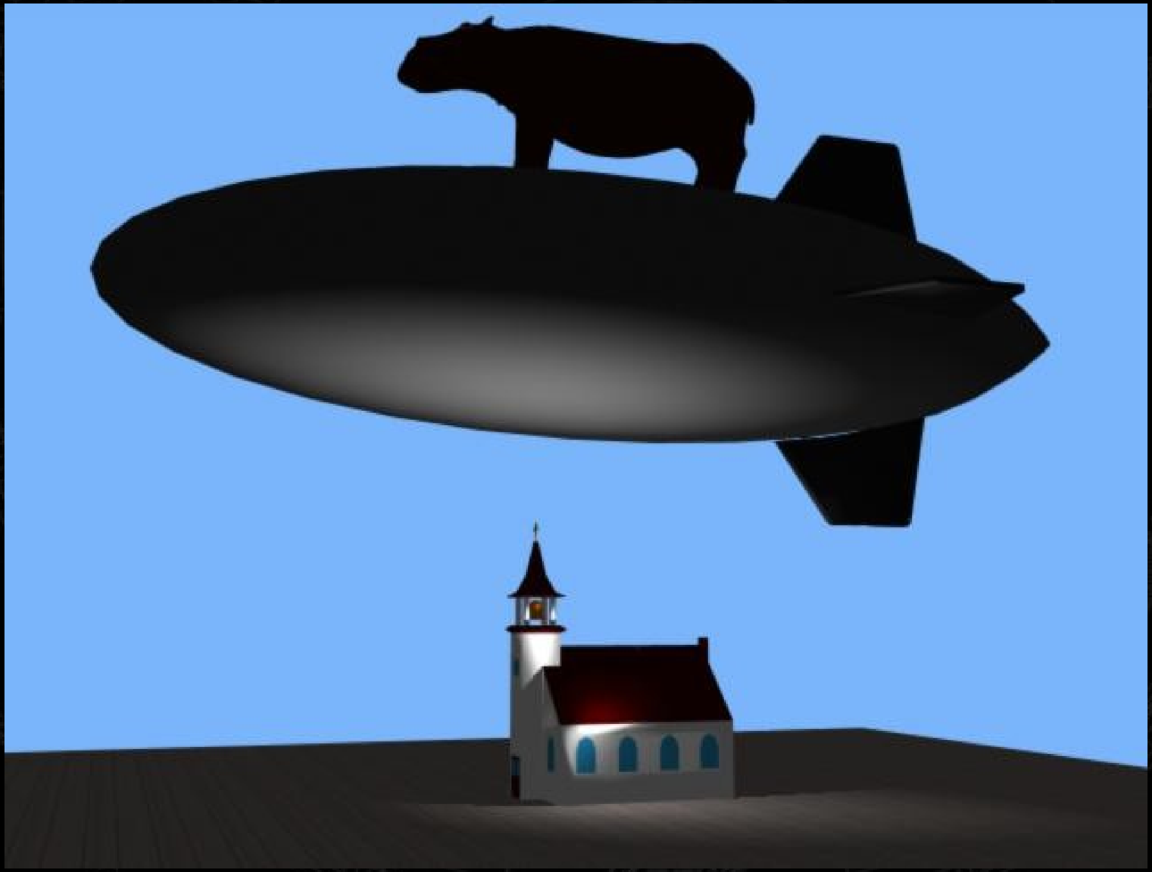
Literalization: *Life is a bowl of cherries.*



Characterization: *The policeman ran by the parking meter*



Functionalization: *The hippo flies over the church*

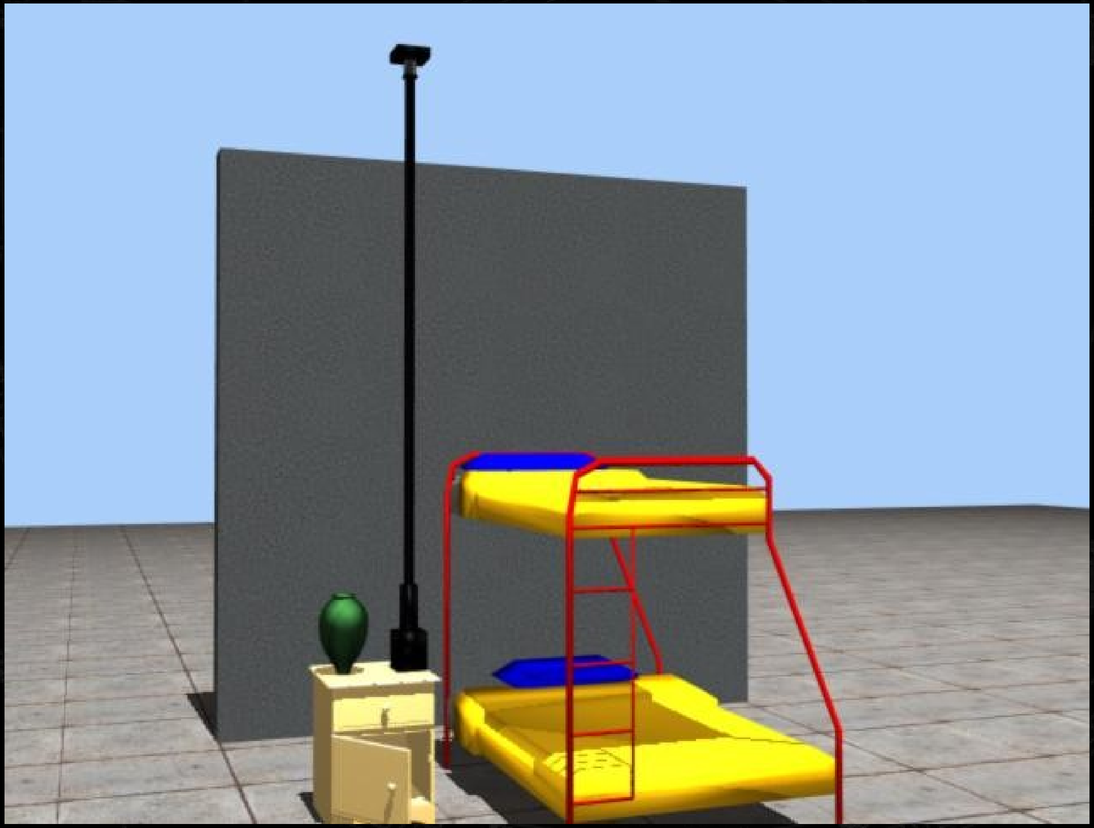


AT&T Labs - Research

Future Work I

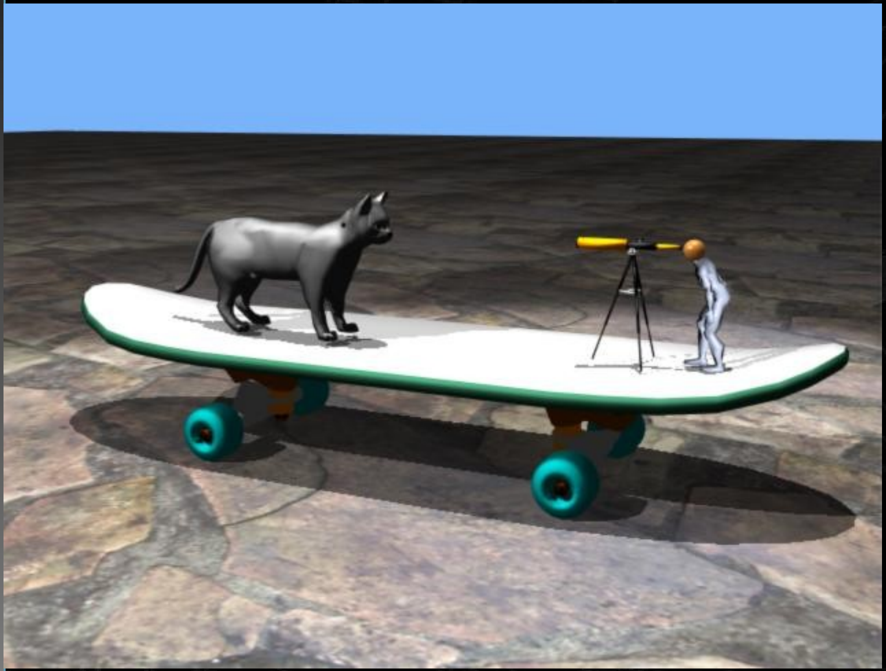
- **Language & Interpretation**
  - Better verb semantics (FrameNet)
  - Exploit world knowledge (Sproat 2001 K-Cap)
  - Compound objects, environments, situations
  - Interactive tweaking
  - Ambiguity issues

Pragmatic Ambiguity: *The lamp is next to the vase on the nightstand…*

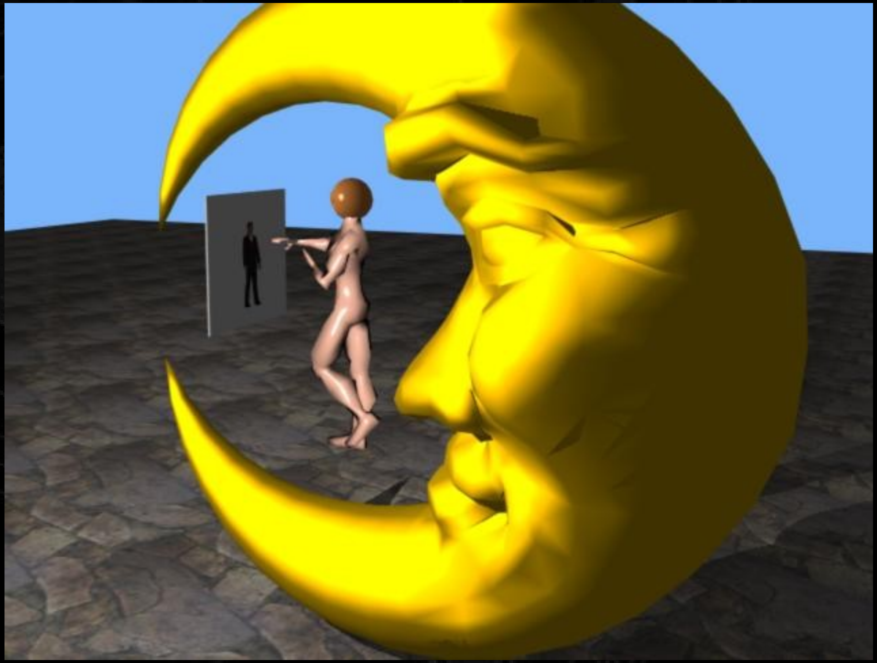


AT&T Labs - Research

Syntactic Ambiguity: Prepositional phrase attachment



John looks at the cat on the skateboard.



John draws the man in the moon.

AT&T Labs - Research

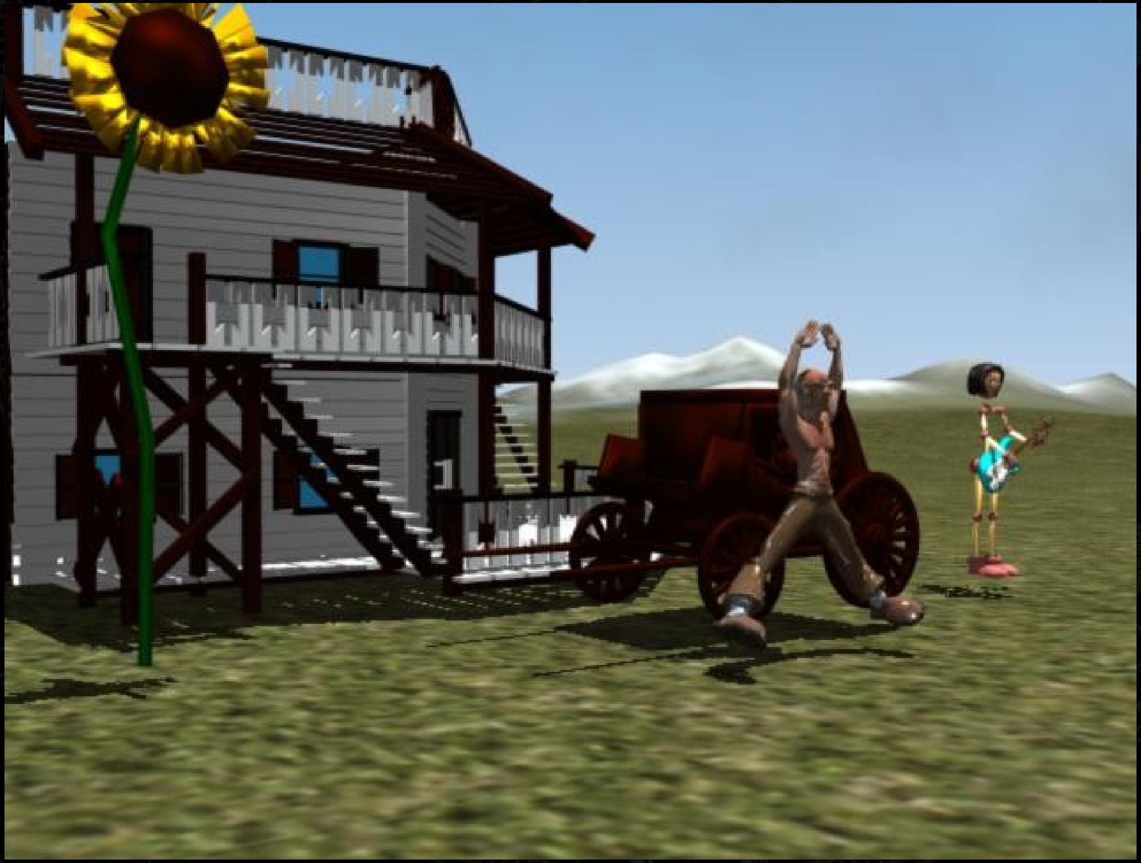
Future Work II

- **Depiction**
  - Handle object parts
  - Improve pose, object, texture DB
  - More complex spatial constraints
  - Dynamic posing, physics
  - Animation

Future Work III

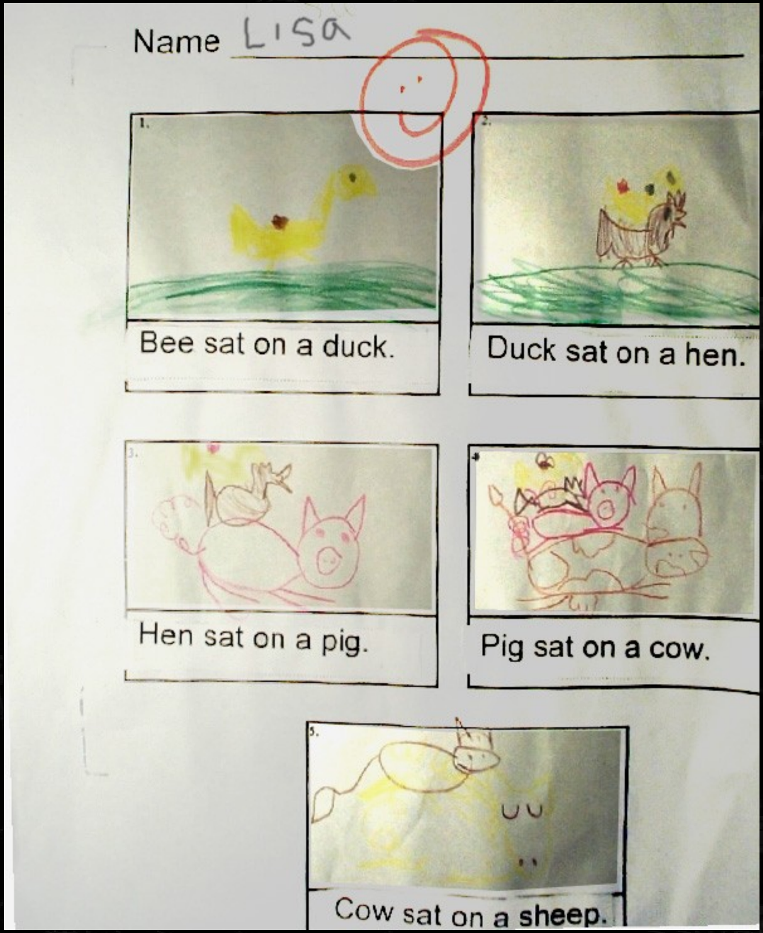
- **Explore Applications**
  - Electronic postcards, visual chat/IM
  - Design (interior, landscape,…)
  - Gaming, virtual environments
  - Storytelling/comic books
  - Art
  - Education

Storytelling: *The stagecoach is in front of the old west hotel. Mary is next to the stagecoach. She plays the guitar. Edward exercises in front of the stagecoach. The large sunflower is to the left of the stagecoach.*



AT&T Labs - Research

1st grade homework: *The duck sat on a hen; the hen sat on a pig;…*



Name Lisa

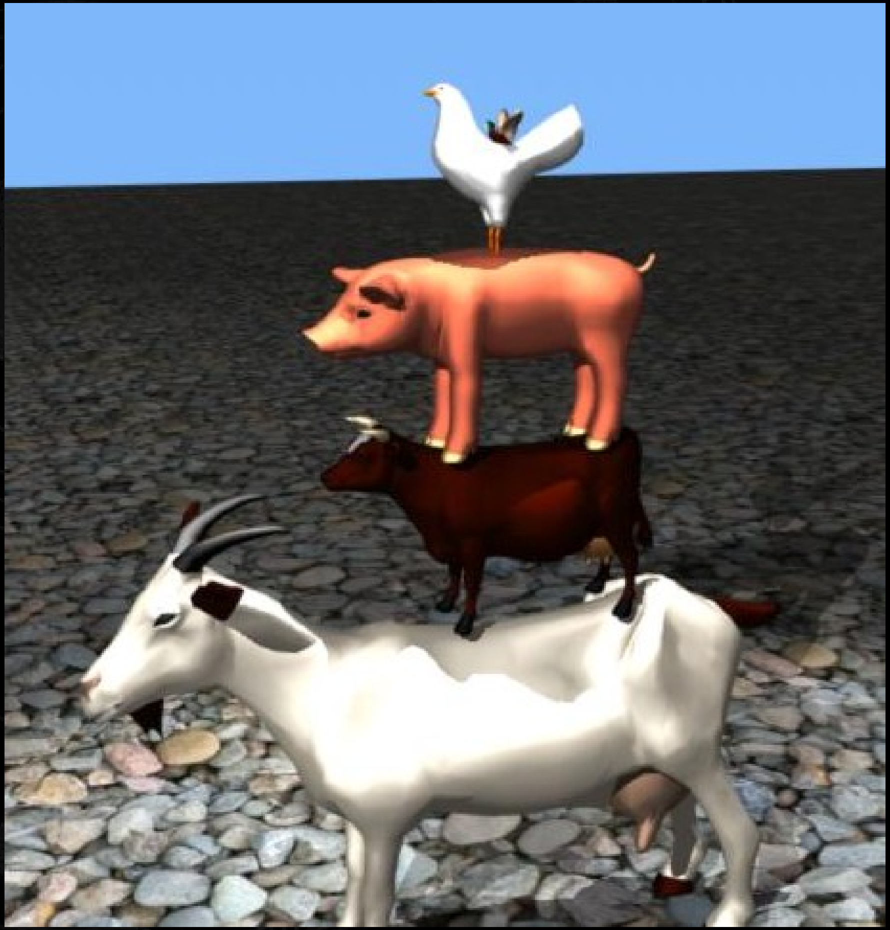
1. Bee sat on a duck.

2. Duck sat on a hen.

3. Hen sat on a pig.

4. Pig sat on a cow.

5. Cow sat on a sheep.



Conclusion

- **New approach to scene generation**
  - Low overhead (skill, training,...)
  - Immediacy
  - Usable with minimal hardware: text or speech input device and display screen.
- **Work is ongoing**
  - User testing of this version by end of year

Acknowledgements

Thanks to:
Adam Buchsbaum, Michael Collins, Martin Kroll, Kirk Mobert, Larry Stead, Gary Zamchick, IZware, audiences at UPenn, AT&T Labs, and Lucent Bell Labs

Web site: www.research.att.com/projects/wordseye